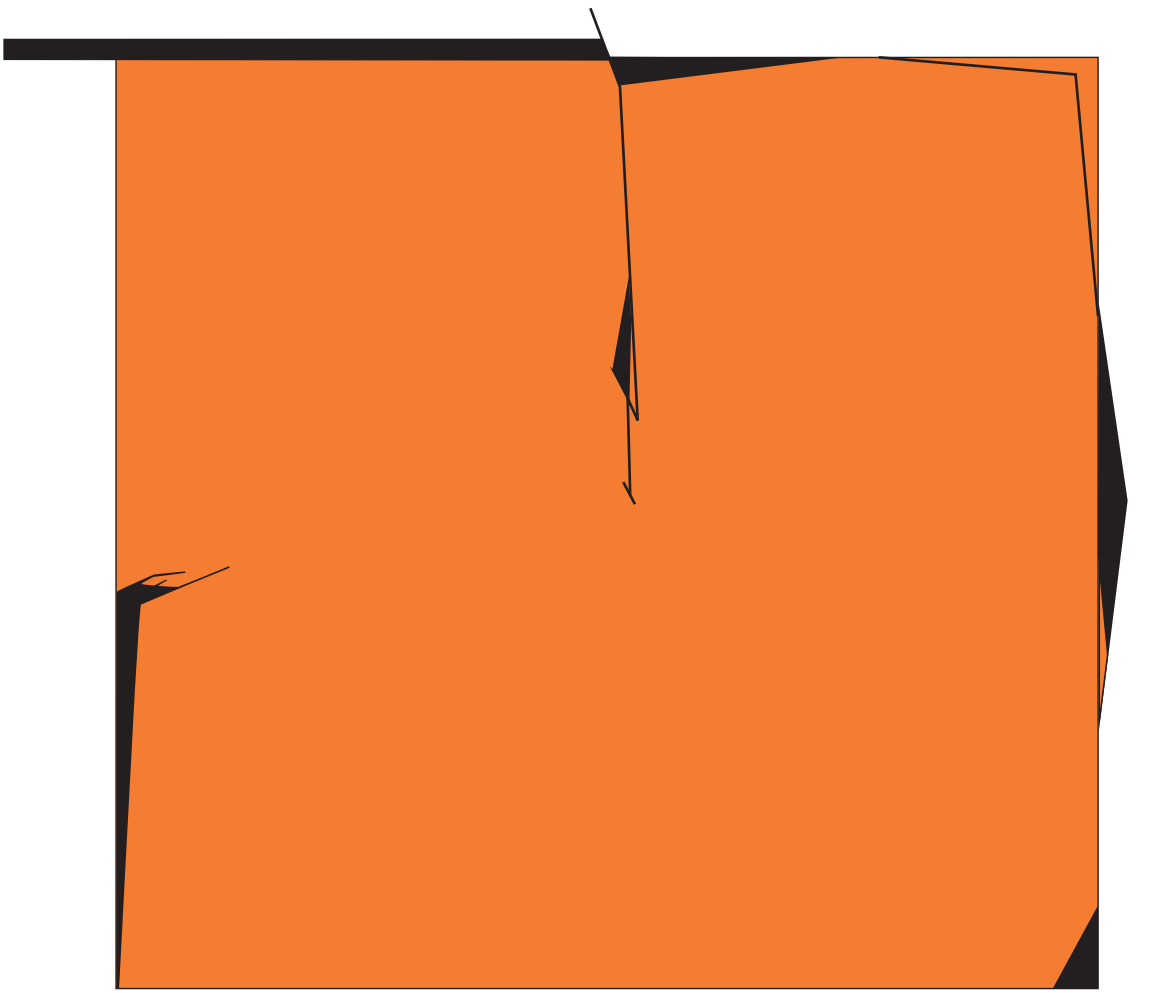
APORTACIONES
MATEMÁTICAS

TEXTOS 39
NIVEL ELEMENTAL

CURSO ELEMENTAL DE PROBABILIDAD Y ESTADÍSTICA

LUIS RINCÓN



SOCIEDAD MATEMÁTICA MEXICANA
2012

CURSO ELEMENTAL DE PROBABILIDAD Y ESTADÍSTICA

Comité Editorial

Marcelo Aguilar González de la Vega *IM, UNAM*
Luz de Teresa de Oteyza *IM, UNAM*
José Ma. González Barrios Murguía *IIMAS, UNAM*
Jesús González Espino Barros *CINVESTAV*
Luis Hernández Lamoneda *CIMAT*
Jorge León Vázquez *CINVESTAV*
Max Neumann Coto *IM, UNAM*
Laura Ortiz Bobadilla *IM, UNAM*
Guillermo Pastor Jiménez *ITAM*
Sergio Rajsbaum Gorodezky *IM, UNAM*
Jorge X. Velasco Hernández *IMP*

Editores Ejecutivos:

Laura Ortiz Bobadilla
Instituto de Matemáticas, UNAM
laura@matem.unam.mx

Guillermo Pastor Jiménez
ITAM
pastor@itam.mx

Instituto de Matemáticas, UNAM
Sociedad Matemática Mexicana

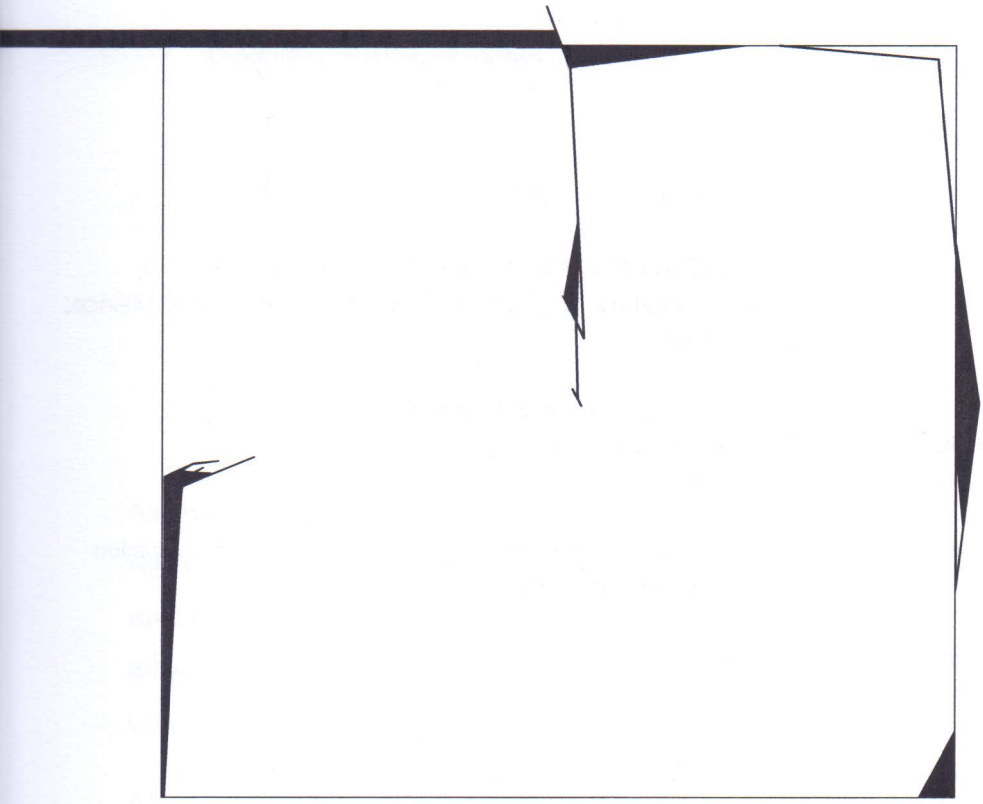
Aportaciones Matemáticas
Textos

APORTACIONES MATEMÁTICAS

TEXTOS 39
NIVEL ELEMENTAL

CURSO ELEMENTAL DE PROBABILIDAD Y ESTADÍSTICA

LUIS RINCÓN



Instituto de Matemáticas
UNAM

Sociedad Matemática Mexicana

México D.F., 2013

Luis Rincón

Departamento de Matemáticas
Facultad de Ciencias
Universidad Nacional Autónoma de México

1a. Edición, 2013

Fecha de edición: 18 de febrero de 2013

D.R. © 2013 UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
Ciudad Universitaria, Delegación Coyoacán, C.P. 04510, México,
Distrito Federal.

ISBN: 978-968-36-3591-4 (Aportaciones Matemáticas)
ISBN: 978-968-36-3594-5 (Serie Textos)
ISBN: 978-607-02-4092-8

Prohibida la reproducción total o parcial por cualquier medio sin la autorización escrita del titular de los derechos patrimoniales.

Impreso y hecho en México.

Edición:
Instituto de Matemáticas, UNAM
Av. Universidad 3000
04510 – México, D.F.

En colaboración con:
Sociedad Matemática Mexicana
Carretera México-Cuernavaca Km 23.5,
Av. Cipreses s/n, San Andrés Totoltepec,
14400 – México, D.F.

Este volumen se imprimió con el apoyo financiero de:
Programa PAPIME Proyecto PE103111, DGAPA, UNAM.

Contenido

Prólogo III

Capítulo 1. Probabilidad 1
1. Introducción 2
2. Probabilidad 8
3. Análisis combinatorio 15
4. Probabilidad condicional e independencia 22
5. Variables aleatorias 29
6. Funciones de densidad y de distribución 32
7. Esperanza, varianza, momentos 40
8. Distribuciones de probabilidad 48
9. Dos teoremas límite 73
10. Simulación 76

Capítulo 2. Estadística 87
1. Introducción 88
2. Variables y tipos de datos 89
3. Estadística descriptiva 91
4. Vectores aleatorios 98
5. Muestras aleatorias y estadísticas 111
6. Estimación puntual 112
7. Estimación por intervalos 118
8. Pruebas de hipótesis 125

Apéndice A. Ejercicios 143

Apéndice B. Solución a los ejercicios 181

Apéndice C. Formulario 211

Bibliografía 219

Índice analítico 221

Prólogo

El presente texto contiene material para un curso semestral de probabilidad y estadística. El objetivo es presentar de manera breve algunas de las ideas fundamentales de estas dos muy útiles ramas de las matemáticas. El texto está dirigido a estudiantes de las distintas carreras de ingeniería y otras carreras científicas similares, cuyos programas de estudio contemplan un curso en donde se muestre el uso y aplicación de la probabilidad y la estadística en sus respectivas áreas de especialidad. Las áreas de aplicación son tan amplias y diversas que no es razonable ni posible intentar mostrarlas todas en un texto como el presente. Se presentan entonces únicamente algunas de las aplicaciones de los resultados teóricos. Debido a la orientación con la que se desea abordar estos temas, no se hace énfasis en el rigor matemático de las demostraciones de los resultados, sino en el uso, interpretación y aplicación de éstos. Sin embargo, para una lectura provechosa de este material, se requiere, en determinados momentos, tener cierta familiaridad con algunos conceptos elementales de álgebra y del cálculo diferencial e integral. A lo largo del texto se presentan numerosos ejemplos y en la parte final se encuentra una colección de ejercicios y algunas sugerencias para resolverlos.

El texto fue escrito en el sistema LaTeX, las gráficas fueron elaboradas con el paquete *PSTricks* y para las ilustraciones se usó el paquete *GIMP*. De acuerdo a la información encontrada en *Wikimedia Commons*, el autor considera que los derechos de autor de las fotos de los personajes que aparecen en este trabajo pertenecen al dominio público, en caso de que esto no sea así para alguna foto se ruega contactar al autor o al editor para omitir la imagen en futuras ediciones o añadir el permiso y crédito correspondientes.

Agradezco sinceramente todos los comentarios que he recibido por parte de alumnos y profesores para mejorar este material, particularmente de los revisores anónimos quienes emitieron dictámenes benevolentes y constructivos acerca de este trabajo. Agradezco también el apoyo del proyecto PAPIME PE103111 a través del cual pudo ser posible el desarrollo y edición de este texto. Estoy muy agradecido también con Nils Ackermann, Gabriela Sanginés y Leonardo Espinosa por su valiosa ayuda en la producción técnica final de este trabajo.

Luis Rincón
Mayo 2012
Ciudad Universitaria UNAM
lars@ciencias.unam.mx

Capítulo 1

Probabilidad

En esta primera mitad del curso estudiaremos algunos conceptos elementales de la teoría matemática de la probabilidad. Esta teoría tuvo como uno de sus primeros puntos de partida el intentar resolver un problema particular concerniente a una apuesta de juego de dados entre dos personas. El problema al que nos referimos involucraba una gran cantidad de dinero y puede plantearse de la siguiente forma:

> *Dos jugadores escogen cada uno de ellos un número del* 1 *al* 6*, distinto uno del otro, y apuestan* 32 *doblones de oro a que el número escogido por uno de ellos aparece en tres ocasiones antes que el número del contrario al lanzar sucesivamente un dado. Suponga que el número de uno de los jugadores ha aparecido dos veces y el número del otro una sola vez. ¿Cómo debe dividirse el total de la apuesta si el juego se suspende?*

Uno de los apostadores, Antonio de Gombaud, popularmente conocido como el caballero De Méré, deseando conocer la respuesta al problema plantea a Blaise Pascal (1623-1662) la situación. Pascal a su vez consulta con Pierre de Fermat (1601-1665) e inician un intercambio de cartas a propósito del problema. Esto sucede en el año de 1654. Con ello se inician algunos esfuerzos por dar solución a éste y otros problemas similares que se plantean. Con el paso del tiempo se sientan las bases y las experiencias necesarias para la búsqueda de una teoría matemática que sintetice los conceptos y los métodos de solución de los muchos problemas particulares resueltos a lo largo de varios años.

En el segundo congreso internacional de matemáticas, celebrado en la ciudad de Paris en el año 1900, el matemático David Hilbert (1862-1943) plantea 23 problemas matemáticos de importancia. Uno de estos problemas es el de encontrar axiomas o postulados a partir de los cuales se pueda construir una teoría matemática de la probabilidad. Aproximadamente treinta años después, en 1933, el matemático ruso A. N. Kolmogorov (1903-1987) propone ciertos axiomas que a la postre resultaron adecuados para la construcción de una teoría de la probabilidad. Esta teoría prevalece hoy en día y ha adquirido el calificativo de teoría clásica. Actualmente la teoría clásica de la probabilidad se ha desarrollado y extendido enormemente gracias a muchos pensadores que han contribuido a su crecimiento, y es sin duda una parte muy importante y bien establecida de las matemáticas. La teoría de la probabilidad ha resultado útil para

Blaise Pascal
(Francia 1623–1662)

Pierre de Fermat
(Francia 1601–1665)

resolver problemas puramente matemáticos, pero sobre todo y principalmente, para modelar situaciones reales o imaginarias, en donde el azar es relevante.

1. Introducción

Existen dos tipos de fenómenos o experimentos en la naturaleza: los deterministas y los aleatorios. Un experimento determinista es aquel que produce el mismo resultado cuando se le repite bajo las mismas condiciones, por ejemplo, medir el volumen de un gas cuando la presión y la temperatura son constantes produce teóricamente siempre el mismo resultado, o medir el ángulo de un rayo de luz reflejado en un espejo resulta siempre en el mismo resultado cuando el ángulo de incidencia es el mismo y el resto de las condiciones son constantes. Muchas otras leyes de la física son ejemplos de situaciones en donde bajo idénticas condiciones iniciales, el resultado del experimento es siempre el mismo. En contraparte, un experimento aleatorio es aquel que cuando se le repite bajo las mismas condiciones, el resultado que se observa no siempre es el mismo y tampoco es predecible. El lanzar una moneda al aire y observar la cara de la moneda que mira hacia arriba, o registrar el número ganador en un juego de lotería, son ejemplos cotidianos de experimentos aleatorios. Nuestro interés en este curso es estudiar ciertos modelos matemáticos, algunas herramientas y varios resultados que los matemáticos y otros pensadores han encontrado y desarrollado en el estudio científico de los experimentos aleatorios. Para ser más precisos, pediremos que los experimentos aleatorios que consideraremos cumplan las siguientes características:

a) El experimento debe poder ser repetible bajo las mismas condiciones iniciales.
b) El resultado de cualquier ensayo del experimento es variable y depende del azar o de algún mecanismo aleatorio.

Es necesario mencionar, sin embargo, que en algunas ocasiones no es fácil clasificar un experimento dado en aleatorio o determinista, esto dependerá del observador, de lo

que él o ella conozca del experimento y de lo que esta persona desea observar en el experimento.

Espacio muestral

Hemos mencionado que la teoría de la probabilidad es la parte de las matemáticas que se encarga del estudio de los fenómenos o experimentos aleatorios. En principio no sabemos cuál será el resultado de un experimento aleatorio, así que por lo menos conviene agrupar en un conjunto a todos los resultados posibles. Esto lleva a la siguiente definición.

DEFINICIÓN 1.1. El espacio muestral, o también llamado espacio muestra, de un experimento aleatorio es el conjunto de todos los posibles resultados del experimento, y se le denota generalmente por la letra griega Ω (omega).

Más adelante mostraremos que el espacio muestral no es necesariamente único y su determinación depende del interés del observador o persona que realiza el experimento aleatorio. En algunos textos se usa también la letra S para denotar al espacio muestral. Esta letra proviene del término *sampling space* de la lengua inglesa, equivalente a espacio muestral. Por otro lado, llamaremos evento o suceso a cualquier subconjunto del espacio muestral y los denotaremos por las primeras letras del alfabeto en mayúsculas: $A, B, C, \ldots$ Con la ayuda de algunos ejemplos ilustraremos a continuación los conceptos de espacio muestral y evento.

EJEMPLO 1.2. Si un experimento aleatorio consiste en lanzar un dado y observar el número que aparece en la cara superior, entonces claramente el espacio muestral es el conjunto $\Omega = \{1, 2, 3, 4, 5, 6\}$. Como ejemplo de un evento para este experimento podemos definir el conjunto $A = \{2, 4, 6\}$, que corresponde al suceso de obtener como resultado un número par. Si al lanzar el dado una vez se obtiene el número "4", decimos entonces que se observó la ocurrencia del evento A, y si se obtiene por ejemplo el resultado "1", decimos que no se observó la ocurrencia del evento A.

EJEMPLO 1.3. Considere el experimento aleatorio de participar en un juego de lotería. Suponga que hay un millón de números en esta lotería y un jugador participa con un boleto. ¿Cuál es un posible espacio muestral para este experimento si únicamente uno de los posibles números es el ganador? Naturalmente al jugador le interesa conocer su suerte en este juego y puede proponer como espacio muestral el conjunto $\Omega =$ {"ganar", "perder" }. Sin embargo puede también tomarse como espacio muestral el conjunto que contiene a todos los posibles números ganadores, es decir, $\Omega = \{1, 2, \ldots, 1000000\}$. Este ejemplo sencillo muestra que el espacio muestral de un experimento aleatorio no es único y depende del interés del observador.

EJEMPLO 1.4. Suponga que un experimento aleatorio consiste en observar el tiempo en el que una máquina en operación sufre su primera descompostura. Si se consideran mediciones continuas del tiempo, entonces puede adoptarse como espacio

muestral para este experimento el intervalo $[0, \infty)$. El subconjunto $A = [1, 2]$ corresponde al evento en el que la primera descompostura se observe entre la primera y la segunda unidad de tiempo.

Diremos que un evento es simple cuando consta de un solo elemento del espacio muestral, en cambio, se llamará compuesto cuando consta de mas de un elemento del espacio muestral. Puesto que los conceptos de espacio muestral y evento involucran forzosamente la terminología de conjuntos, recordaremos a continuación algunas operaciones entre estos objetos y algunas propiedades que nos serán de utilidad en el estudio de la probabilidad y la estadística.

Operaciones con conjuntos

Supondremos que el espacio muestral Ω de un experimento aleatorio es una especie de conjunto universal y cualquier elemento de Ω lo denotaremos por ω (omega minúscula). Al conjunto vacío lo denotaremos como es usual por el símbolo $\emptyset$. Otros símbolos usuales son los de pertenencia ($\in$), o no pertenencia ($\notin$) de un elemento en un conjunto, y los de contención ($\subset, \subseteq$), o no contención ($\not\subset$) de un conjunto en otro. ¿Puede usted explicar el significado del símbolo $\subsetneq$? Justamente, decimos que A es un subconjunto propio de B si $A \subsetneq B$. La igualdad de dos conjuntos A y B significa que se cumplen las dos contenciones: $A \subset B$ y $B \subset A$. Por último, si A es un conjunto, denotamos la cardinalidad o número de elementos de ese conjunto por el símbolo $\#A$.

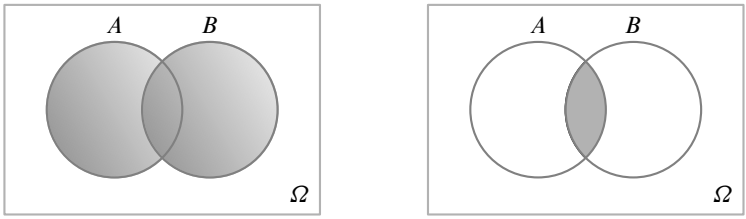
Sean A y B dos subconjuntos cualesquiera de Ω. Recordamos a continuación las operaciones básicas de unión, intersección, diferencia y complemento:

$$\begin{aligned} A \cup B &= \{\omega \in \Omega : \omega \in A \text{ ó } \omega \in B\}, \\ A \cap B &= \{\omega \in \Omega : \omega \in A \text{ y } \omega \in B\}, \\ A - B &= \{\omega \in \Omega : \omega \in A \text{ y } \omega \notin B\}, \\ A^c &= \{\omega \in \Omega : \omega \notin A\}. \end{aligned}$$

Cuando los conjuntos se expresan en palabras, la operación unión, $A \cup B$, se lee "A ó B" y la intersección, $A \cap B$, se lee "A y B". En la Figura 1.1 se muestran en diagramas de Venn estas dos operaciones.

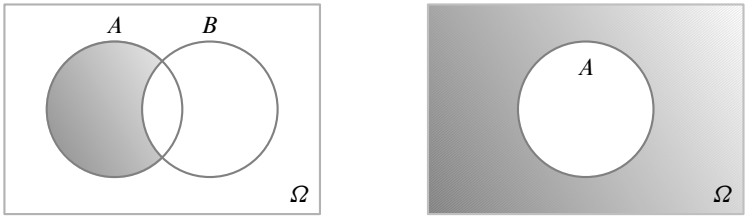
La diferencia entre dos conjuntos A y B se denota por $A - B$, y corresponde a aquel conjunto de elementos de A que no pertenecen a B, es decir, $A - B$ se define como $A \cap B^c$. En general, el conjunto $A - B$ es distinto de $B - A$, de hecho estos conjuntos son siempre ajenos, ¿puede usted comprobar tal afirmación? ¿en qué caso ambos conjuntos coinciden? Por otro lado el complemento de un conjunto A se denota por A^c y se define como la colección de aquellos elementos de Ω que no pertenecen a A. Mediante un diagrama de Venn ilustramos gráficamente las operaciones de diferencia y complemento en la Figura 1.2.

EJEMPLO 1.5. Sea A el conjunto de aquellas personas que tienen hijos y B la colección de aquellas personas que están casadas. Entonces el conjunto $A \cap B$ consta



$A \cup B$ $A \cap B$

FIGURA 1.1



$A - B$ A^c

FIGURA 1.2

de aquellas personas que están casadas y tienen hijos, mientras que el conjunto $A \cap B^c$ está constituido por aquellas personas que tienen hijos pero no están casadas. ¿Quién es $A^c \cap B$? Observe que cada persona es un elemento de alguno de los siguientes conjuntos: $A \cap B$, $A \cap B^c$, $A^c \cap B$ ó $A^c \cap B^c$. ¿A cuál pertenece usted?

Es fácil verificar que el conjunto vacío y el conjunto total satisfacen las siguientes propiedades elementales: $A \cup \emptyset = A$, $A \cap \Omega = A$, $A \cap \emptyset = \emptyset$, $A \cup A^c = \Omega$, $A \cup \Omega = \Omega$, $A \cap A^c = \emptyset$. Además las operaciones unión e intersección son asociativas, esto es, satisfacen las siguientes igualdades:

$$\begin{aligned} A \cup (B \cup C) &= (A \cup B) \cup C, \\ A \cap (B \cap C) &= (A \cap B) \cap C, \end{aligned}$$

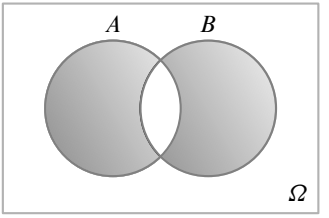
y también son distributivas, es decir,

$$\begin{aligned} A \cap (B \cup C) &= (A \cap B) \cup (A \cap C), \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C). \end{aligned}$$

Recordemos también la operación diferencia simétrica entre dos conjuntos A y B, denotada por $A \triangle B$, y definida como sigue

$$A \triangle B = (A \cup B) - (B \cap A).$$

En la Figura 1.3 ilustramos gráficamente el conjunto resultante de efectuar la diferencia simétrica entre los conjuntos A y B. Visualmente es fácil comprobar que la diferencia simétrica también puede escribirse como $(A - B) \cup (B - A)$. ¿Cómo podría expresarse en palabras al conjunto $A \triangle B$?



$A \triangle B$

FIGURA 1.3

Recordemos además las muy útiles leyes de De Morgan:

$$\begin{aligned} (A \cup B)^c &= A^c \cap B^c, \\ (A \cap B)^c &= A^c \cup B^c. \end{aligned}$$

La validez de estas dos igualdades puede extenderse a colecciones finitas e incluso arbitrarias de conjuntos. ¿Puede usted escribir estas identidades para n conjuntos?

Conjuntos ajenos

Cuando dos conjuntos no tienen ningún elemento en común se dice que son ajenos, la definición formal es la siguiente.

DEFINICIÓN 1.6. Decimos que dos conjuntos A y B son ajenos o disjuntos si se cumple la igualdad

$$A \cap B = \emptyset.$$

Es decir, los conjuntos A y B son ajenos cuando no existe un elemento que pertenezca tanto a A como a B. Por ejemplo, si $\Omega = \{1, 2, 3, 4, 5, 6\}$, entonces los conjuntos $A = \{1, 2\}$ y $B = \{3, 4\}$ son ajenos pues no hay ningún elemento común entre ellos. El ejemplo general más importante de conjuntos o eventos ajenos es la pareja dada por A y A^c, para cualquier conjunto A. La propiedad de ser ajenos puede extenderse al caso cuando se tienen varios conjuntos. Decimos que n conjuntos $A_1, \ldots, A_n$ son ajenos si $A_1 \cap \cdots \cap A_n = \emptyset$, y se dice que son ajenos dos a dos (o mutuamente ajenos) si $A_i \cap A_j = \emptyset$ para cualesquiera valores de los índices $i, j = 1, 2, \ldots, n$, con i distinto de j. Existe diferencia en estas dos definiciones y explicaremos la situación con el siguiente ejemplo.

EJEMPLO 1.7. Los conjuntos $A = \{1, 2\}$, $B = \{2, 3\}$ y $C = \{3, 4\}$ son ajenos pues $A \cap B \cap C = \emptyset$, pero no son ajenos dos a dos pues, por ejemplo, el conjunto $A \cap B$ no es vacío. Así, los conjuntos A, B y C son ajenos en el sentido de que la intersección de todos ellos es vacía pero no son ajenos dos a dos.

Las operaciones entre conjuntos que mencionaremos a continuación no son elementales y producen nuevos conjuntos que se encuentran en un nivel distinto al de los conjuntos originales.

Conjunto potencia

El conjunto potencia de Ω, denotado por 2^Ω, es aquel conjunto cuyos elementos son todos los subconjuntos posibles de Ω. En términos estrictos esta nueva colección deja de ser un conjunto y se le llama clase de subconjuntos de Ω, aunque seguiremos usando el primer término en nuestro tratamiento elemental de los conjuntos. Por ejemplo, si $\Omega = \{a, b, c\}$, entonces el conjunto 2^Ω consta de 8 elementos, a saber,

$$2^\Omega = \Big\{\emptyset, \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\Big\}.$$

Observe que los elementos del conjunto potencia son en sí mismos conjuntos, y que en esta colección están contenidos todos los eventos que podrían ser de interés en un experimento aleatorio. No es difícil demostrar que

$$\#(2^\Omega) = 2^{\#\Omega},$$

es decir, el número de elementos en el conjunto 2^Ω es exactamente 2 elevado a la potencia dada por la cardinalidad de Ω. De este hecho proviene la notación usada para el conjunto potencia: 2^Ω. Observe que la expresión 2^Ω no tiene sentido matemático, y debe considerarse como un símbolo para denotar al conjunto potencia. Para el ejemplo anterior se comprueba que la cardinalidad de 2^Ω es efectivamente $2^{\#\Omega} = 2^3 = 8$.

Producto Cartesiano

Finalmente recordemos que el producto Cartesiano de dos conjuntos A y B, denotado por $A \times B$, se define como la colección de todas las parejas ordenadas (a, b), en donde a es cualquier elemento de A, y b es cualquier elemento de B. En símbolos,

$$A \times B = \{(a, b) : a \in A \text{ y } b \in B\}.$$

Por ejemplo, si $A = \{a_1, a_2\}$ y $B = \{b_1, b_2, b_3\}$, entonces

$$A \times B = \{(a_1, b_1), (a_1, b_2), (a_1, b_3), (a_2, b_1), (a_2, b_2), (a_2, b_3)\}.$$

Este conjunto puede representarse gráficamente mediante un diagrama de árbol como el que se ilustra en la Figura 1.4 .

En general los conjuntos producto $A \times B$ y $B \times A$ son distintos pues la pareja (a, b) es distinta de (b, a), sin embargo ambos conjuntos tienen la misma cardinalidad, esto es, ambos tienen el mismo número de elementos. Más aún, si la cardinalidad de

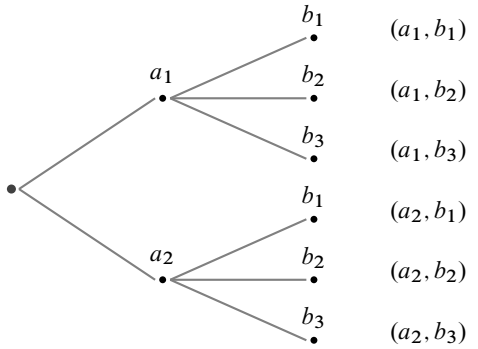


FIGURA 1.4. Diagrama de árbol.

A es el número n, y la cardinalidad de B es m, entonces la cardinalidad del conjunto $A \times B$ es el producto $n \cdot m$. Este resultado es llamado principio de multiplicación y lo aplicaremos con frecuencia más adelante. Un poco más generalmente, puede considerarse el producto Cartesiano de n conjuntos y comprobarse que

$$\#(A_1 \times \cdots \times A_n) = \#A_1 \cdots \#A_n.$$

EJEMPLO 1.8. Considere el producto Cartesiano $\mathbb{R} \times \mathbb{R}$, que es el conjunto de todas las parejas de números reales (x, y). A este conjunto producto se le denota usualmente por $\mathbb{R}^2$. Análogamente se construyen los conjuntos $\mathbb{R}^3$, $\mathbb{R}^4$, ..., $\mathbb{R}^n$.

Concluimos aquí nuestra rápida y breve revisión de la teoría elemental de conjuntos. Recordemos que estamos interesados en calcular probabilidades de los diferentes eventos, es decir, de subconjuntos del espacio muestral que se obtienen al estudiar los diversos experimentos aleatorios. En la siguiente sección estudiaremos algunas formas de definir matemáticamente la probabilidad de un evento.

2. Probabilidad

La probabilidad de un evento A, es un número real en el intervalo $[0, 1]$ que denotaremos por $P(A)$, y representa una medida de la frecuencia con la que se observa la ocurrencia del evento A cuando se efectúa el experimento aleatorio en cuestión. Existen al menos cuatro definiciones de probabilidad, las cuales explicamos a continuación.

Probabilidad clásica

Sea A un subconjunto de un espacio muestral Ω de cardinalidad finita. Se define la probabilidad clásica del evento A como el cociente:

$$P(A) = \frac{\#A}{\#\Omega},$$

en donde el símbolo #A denota la cardinalidad o número de elementos del conjunto A. Claramente esta definición es sólo válida para espacios muestrales finitos, pues forzosamente necesitamos suponer que el número de elementos en Ω es finito. Además, el espacio Ω debe ser equiprobable, pues para calcular la probabilidad de un evento A, únicamente necesitamos contar cuántos elementos tiene A respecto del total Ω, sin importar exactamente qué elementos particulares sean. Por lo tanto, esta definición de probabilidad presupone que todos los elementos de Ω son igualmente probables o tienen el mismo peso. Este es el caso por ejemplo de un dado equilibrado. Para este experimento el espacio muestral es el conjunto $\Omega = \{1, 2, 3, 4, 5, 6\}$, y si deseamos calcular la probabilidad (clásica) del evento A correspondiente a obtener un número par, es decir, la probabilidad de $A = \{2, 4, 6\}$, entonces

$$P(A) = \frac{\#\{2, 4, 6\}}{\#\{1, 2, 3, 4, 5, 6\}} = \frac{3}{6} = \frac{1}{2}.$$

A esta forma definir la probabilidad también se le conoce con el nombre de probabilidad de Laplace, en honor del astrónomo y matemático francés Pierre-Simon Laplace, quien estableció de una manera sistemática y rigurosa, los principios y propiedades de esta manera de calcular probabilidades.

Pierre-Simon Laplace
(Francia 1749–1827)

Probabilidad frecuentista

Suponga que se realizan n repeticiones de un cierto experimento aleatorio y sea A un evento cualquiera. Denotemos por $n(A)$ el número de ocurrencias del evento A en las n realizaciones del experimento. Se define entonces la probabilidad frecuentista de A como indica el siguiente límite

$$P(A) = \lim_{n\to\infty} \frac{n(A)}{n}.$$

En este caso, debemos hacer notar que no es humanamente posible llevar a cabo una infinidad de veces el experimento aleatorio, de modo que en la práctica no es posible encontrar mediante este mecanismo la probabilidad de un evento cualquiera. Esta limitación hace que esta definición de probabilidad no sea enteramente formal, pero tiene algunas ventajas. Veamos un ejemplo concreto. Consideremos nuevamente el experimento aleatorio de lanzar un dado equilibrado y registrar la ocurrencia del evento A definido como el conjunto $\{2, 4, 6\}$. Después de lanzar el dado 20 veces se obtuvieron los siguientes resultados:

No.	Resultado	$n(A)/n$
1	3	0/1
2	6	1/2
3	2	2/3
4	1	2/4
5	4	3/5
6	6	4/6
7	3	4/7
8	4	5/8
9	2	6/9
10	5	6/10

No.	Resultado	$n(A)/n$
11	2	7/11
12	5	7/12
13	1	7/13
14	6	8/14
15	3	8/15
16	1	8/16
17	5	8/17
18	5	8/18
19	2	9/19
20	6	10/20

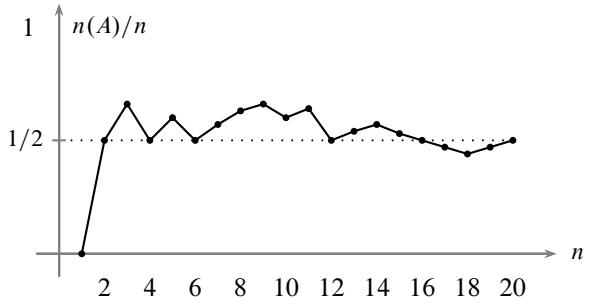


FIGURA 1.5

En la gráfica de la Figura 1.5 se muestra el singular comportamiento de este cociente a lo largo del tiempo, al principio se pueden presentar algunas oscilaciones pero eventualmente el cociente se estabiliza en un cierto número. Realizando un mayor número de observaciones del experimento, no es difícil creer que el cociente $n(A)/n$ se estabiliza en $1/2$ cuando n es grande y suponiendo que el dado esta equilibrado. Se invita al lector intrigado a efectuar un experimento similar y corroborar esta interesante regularidad estadística con éste o cualquier otro experimento aleatorio de su interés.

Más adelante formalizaremos este resultado mediante la así llamada ley de los grandes números.

SIMULACIÓN 1.9. Arroje cien veces una moneda y registre los resultados en una lista. Calcule y grafique los cocientes $n(A)/n$ cuando el evento A corresponde a obtener alguna de las caras de la moneda. ¿Converge el cociente $n(A)/n$ a $1/2$? Para agilizar el experimento puede usted dividir los cien lanzamientos en varios grupos de personas y después juntar los resultados. Alternativamente, investigue la forma de simular este experimento en una computadora y calcule $n(A)/n$ para distintos valores de n.

Probabilidad subjetiva

En este caso la probabilidad de un evento depende del observador, es decir, depende de lo que el observador conoce del fenómeno en estudio. Puede parecer un tanto informal y poco seria esta definición de la probabilidad de un evento, sin embargo en muchas situaciones es necesario recurrir a un experto para tener por lo menos una idea vaga de cómo se comporta el fenómeno de nuestro interés y saber si la probabilidad de un evento es alta o baja. Por ejemplo, ¿cuál es la probabilidad de que un cierto equipo de fútbol gane en su próximo partido? Ciertas circunstancias internas del equipo, las condiciones del equipo rival o cualquier otra condición externa, son elementos que sólo algunas personas conocen y que podrían darnos una idea más exacta de esta probabilidad. Esta forma subjetiva de asignar probabilidades a los distintos eventos debe, sin embargo, ser consistente con una serie de reglas naturales que estudiaremos a continuación.

Probabilidad axiomática

En la definición axiomática de la probabilidad no se establece la forma explícita de calcular las probabilidades sino únicamente se proponen las reglas que el cálculo de probabilidades debe satisfacer. Los siguientes tres postulados o axiomas[1] fueron establecidos en 1933 por el matemático ruso Andrey Nikolaevich Kolmogorov.

Axiomas de la probabilidad

1. $P(A) \geq 0$.
2. $P(\Omega) = 1$.
3. $P(A \cup B) = P(A) + P(B)$, cuando $A \cap B = \emptyset$.

[1] Un postulado o axioma es una proposición que se acepta como válida y sobre la cual se funda una teoría.

No es difícil verificar que las definiciones anteriores de probabilidad satisfacen estos tres axiomas. De hecho, estos postulados han sido tomados directamente del análisis cuidadoso y reflexivo de las definiciones de probabilidad mencionadas anteriormente. Por ejemplo, considerando nuevamente la definición de probabilidad frecuentista, se realiza una sucesión de n ensayos de un experimento aleatorio y se calcula el cociente $n(A)/n$ para un evento A cualquiera. Se observa claramente que el cociente $n(A)/n$ es no negativo, esto es el primer axioma. Si A es el evento total Ω, entonces $n(\Omega) = n$ y por lo tanto $n(\Omega)/n = 1$, esto es el segundo axioma. Finalmente, si A y B son dos eventos ajenos, se tiene que $n(A \cup B) = n(A) + n(B)$ y por lo tanto

$$\frac{n(A \cup B)}{n} = \frac{n(A)}{n} + \frac{n(B)}{n},$$

de donde surge el tercer axioma. Observe además que este tercer axioma es válido no sólo para dos eventos ajenos sino para cualquier colección finita de eventos ajenos dos a dos. A cualquier función P que satisfaga los tres axiomas de Kolmogorov se le llama medida de probabilidad, o simplemente probabilidad. A partir de estos postulados es posible demostrar que la probabilidad cumple con una serie de propiedades interesantes.

PROPOSICIÓN 1.10. *Para cualquier evento A,* $P(A^c) = 1 - P(A)$.

DEMOSTRACIÓN. De la teoría elemental de conjuntos tenemos que $\Omega = A \cup A^c$. Como A y A^c son eventos ajenos, por el tercer axioma, $P(\Omega) = P(A) + P(A^c)$. Finalmente, como $P(\Omega) = 1$, por el segundo axioma obtenemos $P(A^c) = 1 - P(A)$. ■

La proposición recién demostrada establece que los eventos A y A^c tienen probabilidad complementaria, es decir, la suma de las probabilidades de estos eventos es siempre uno. Esta propiedad es bastante útil pues en ocasiones es más sencillo calcular la probabilidad del complemento de un evento que del evento mismo. Más adelante tendremos múltiples ocasiones para aplicar este resultado.

PROPOSICIÓN 1.11. $P(\emptyset) = 0$.

DEMOSTRACIÓN. Como $\emptyset = \Omega^c$, usando la propiedad anterior, tenemos que $P(\emptyset) = P(\Omega^c) = 1 - P(\Omega) = 0$. ■

Las siguientes dos proposiciones suponen la situación $A \subseteq B$ que se muestra gráficamente en la Figura 1.6.

PROPOSICIÓN 1.12. *Si $A \subseteq B$, entonces* $P(A) \leq P(B)$.

DEMOSTRACIÓN. Primeramente escribimos $B = A \cup (B - A)$. Como A y $B - A$ son eventos ajenos, por el tercer axioma, $P(B) = P(A) + P(B - A)$. Usando el primer axioma concluimos que $P(B) - P(A) = P(B - A) \geq 0$. De aquí obtenemos $P(B) - P(A) \geq 0$. ■

PROPOSICIÓN 1.13. *Si $A \subseteq B$, entonces* $P(B - A) = P(B) - P(A)$.

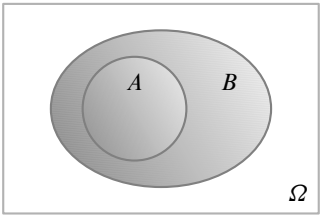


FIGURA 1.6. $A \subseteq B$.

DEMOSTRACIÓN. Como $B = A \cup (B - A)$, siendo esta unión ajena, por el tercer axioma tenemos que $P(B) = P(A) + P(B - A)$. ■

Por ejemplo, suponga que A y B son eventos tales que $A \subseteq B$, $P(A^c) = 0.9$ y $P(B^c) = 0.6$. Deseamos calcular $P(B - A)$. En esta situación es válida la fórmula $P(B - A) = P(B) - P(A)$, en donde, $P(A) = 0.1$ y $P(B) = 0.4$. Por lo tanto, $P(B - A) = 0.4 - 0.1 = 0.3$. Observe que en este ejemplo sencillo no se especifica el experimento aleatorio en cuestión ni tampoco se definen los eventos A y B. El tratamiento es analítico y los resultados son válidos para cualesquiera eventos A y B con las características señaladas.

PROPOSICIÓN 1.14. *Para cualquier evento A,* $0 \leq P(A) \leq 1$.

DEMOSTRACIÓN. Como $A \subseteq \Omega$, se tiene que $P(A) \leq P(\Omega) = 1$. La otra desigualdad, $0 \leq P(A)$, es simplemente el primer axioma. ■

PROPOSICIÓN 1.15. *Para cualesquiera eventos A y B,*

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

DEMOSTRACIÓN. Primeramente observamos que para cualesquiera eventos A y B se cumple la igualdad $A - B = A - (A \cap B)$. Entonces escribimos a $A \cup B$ como la unión disjunta de los siguientes tres eventos:

$$\begin{aligned} A \cup B &= (A - B) \cup (A \cap B) \cup (B - A) \\ &= (A - A \cap B) \cup (A \cap B) \cup (B - A \cap B). \end{aligned}$$

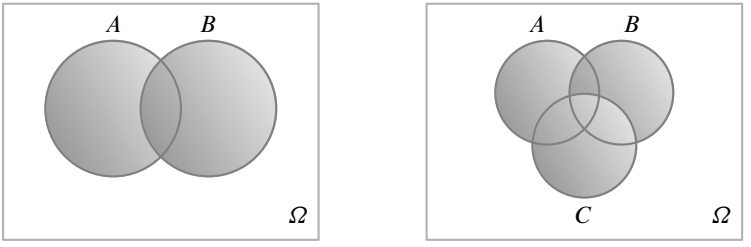
Ahora aplicamos la probabilidad. Por el tercer axioma,

$$P(A \cup B) = P(A - A \cap B) + P(A \cap B) + P(B - A \cap B).$$

Pero $A \cap B \subseteq A$, de modo que $P(A - A \cap B) = P(A) - P(A \cap B)$. Análogamente $P(B - A \cap B) = P(B) - P(A \cap B)$. Por lo tanto,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$ ■

En la Figura 1.7 (a) el lector puede comprobar la validez de la fórmula anterior identificando las tres regiones ajenas de las que consta $A \cup B$. El término $P(A)$ abarca las primeras dos regiones de izquierda a derecha, $P(B)$ abarca la segunda y tercera región. Observe entonces que la región central ha sido contada dos veces de modo que el término $-P(A \cap B)$ da cuenta de ello. De esta forma las tres regiones son contadas una sola vez y el resultado es la probabilidad del evento $A \cup B$.



(a) $A \cup B$ (b) $A \cup B \cup C$

FIGURA 1.7

EJEMPLO 1.16. Sean A y B eventos ajenos tales que $P(B) = 0.3$ y $P(A \cap B^c) = 0.2$.Encuentre $P(A \cup B)$.
Solución: Usaremos la fórmula $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Conocemos $P(B)$. Además $P(A \cap B)$ es cero pues por hipótesis los eventos son ajenos, y $P(A) = P(A \cap B^c) = 0.2$. ¿Por qué? Por lo tanto, $P(A \cup B) = 0.2 + 0.3 = 0.5$.

Observe que la fórmula anterior es válida para cualesquiera eventos A y B. En particular, cuando son conjuntos ajenos, es decir, cuando $A \cap B = \emptyset$, entonces la fórmula demostrada se reduce al tercer axioma de la probabilidad, es decir, $P(A \cup B) = P(A) + P(B)$. El siguiente resultado es una generalización del anterior e involucra tres eventos cualesquiera. La fórmula que a continuación se demuestra puede también verificarse usando el diagrama de Venn que aparece en la Fig 1.7 (b). Para ello siga los términos del lado derecho de la fórmula y compruebe que cada región es contada una sola vez de modo que el resultado final es la probabilidad del evento $A \cup B \cup C$.

PROPOSICIÓN 1.17. *Para cualesquiera eventos A, B y C,*

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) \\ &\quad - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C). \end{aligned}$$

DEMOSTRACIÓN. Usando la fórmula para dos eventos y agrupando adecuadamente,

$$\begin{aligned} P(A \cup B \cup C) &= P[(A \cup B) \cup C] \\ &= P(A \cup B) + P(C) - P((A \cup B) \cap C) \\ &= P(A) + P(B) - P(A \cap B) + P(C) \\ &\quad -P((A \cap C) \cup (B \cap C)) \\ &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) \\ &\quad -P(B \cap C) + P(A \cap B \cap C). \end{aligned}$$

■

Las propiedades anteriores son parte del estudio teórico y general de la probabilidad. En general, supondremos que la forma explícita de calcular estos números es conocida, o que se puede suponer un cierto modelo para llevar a cabo estos cálculos dependiendo del experimento aleatorio en cuestión. Por ejemplo, cuando el espacio muestral es finito y cada resultado puede suponerse igualmente probable, entonces usaremos la definición clásica de probabilidad. En otras situaciones asignaremos probabilidades de acuerdo a ciertos modelos conocidos. Regresaremos a este punto más adelante. A manera de conclusión presentamos a continuación un resumen con las propiedades de la probabilidad que hemos demostrado.

a) $P(A^c) = 1 - P(A)$.
b) $P(\emptyset) = 0$.
c) Si $A \subseteq B$, entonces $P(A) \leq P(B)$.
d) Si $A \subseteq B$, entonces $P(B - A) = P(B) - P(A)$.
e) $0 \leq P(A) \leq 1$.
f) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
g) $P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$.

Esperamos que, a partir de las propiedades enunciadas y demostradas, el lector haya desarrollado cierta habilidad e intuición para escribir la demostración de alguna otra propiedad de la probabilidad. Otras propiedades sencillas de la probabilidad pueden encontrarse en la sección de ejercicios. Debemos también mencionar que las demostraciones no son únicas y que es altamente probable que el lector pueda producir alguna demostración diferente a las que aquí se han presentado.

3. Análisis combinatorio

Consideraremos ahora el caso cuando el experimento aleatorio es tal que su espacio muestral es un conjunto finito y cada elemento de este conjunto tiene la misma probabilidad de ocurrir, es decir, cuando el espacio Ω es finito y equiprobable. En estos casos hemos definido la probabilidad clásica de un evento A de la siguiente forma:

$P(A) = \#A/\#\Omega$. Para poder aplicar esta definición necesitamos saber contar cuántos elementos tiene un evento A cualquiera. Cuando podemos poner en una lista todos y cada uno de los elementos de dicho conjunto, entonces es fácil conocer la cardinalidad de A, simplemente contamos todos los elementos uno por uno. Sin embargo, es común enfrentar situaciones en donde no es factible escribir en una lista cada elemento de A, por ejemplo, ¿cuántos números telefónicos existen que contengan por lo menos un cinco? Es poco factible que alguien intente escribir uno a uno todos estos números telefónicos y encuentre de esta manera la cantidad buscada. En las siguientes secciones estudiaremos algunas técnicas de conteo que nos ayudarán a calcular la cardinalidad de un evento A en ciertos casos particulares. El principio de multiplicación que enunciamos a continuación es la base de muchos de los cálculos en las técnicas de conteo.

PROPOSICIÓN 1.18 (Principio de multiplicación). *Si un procedimiento A_1 puede efectuarse de n formas distintas y un segundo procedimiento A_2 puede realizarse de m formas diferentes, entonces el total de formas en que puede efectuarse el primer procedimiento seguido del segundo es el producto $n \cdot m$, es decir,* $\#(A_1 \times A_2) = \#A_1 \cdot \#A_2$.

Para ilustrar este resultado considere el siguiente ejemplo: suponga que un cierto experimento aleatorio consiste en lanzar un dado y después seleccionar al azar una letra del alfabeto. ¿Cuál es la cardinalidad del correspondiente espacio muestral? El experimento de lanzar un dado tiene 6 resultados posibles y consideremos que tenemos un alfabeto de 26 letras. El correspondiente espacio muestral tiene entonces cardinalidad $6 \times 26 = 156$. El principio de multiplicación es válido no solamente para dos procedimientos sino que también vale para cualquier sucesión finita de procedimientos. Por ejemplo, si $A_1, A_2, \ldots, A_k$ denotan k procedimientos sucesivos, entonces este principio se puede enunciar en símbolos de la forma siguiente: $\#(A_1 \times \cdots \times A_k) = \#A_1 \cdots \#A_k$.

EJEMPLO 1.19. Un hombre tiene 4 pantalones distintos, 6 camisas, y dos pares de zapatos. ¿De cuántas formas distintas puede el hombre vestirse con estas prendas? Solución: $4 \times 6 \times 2 = 48$. Es decir, el hombre se puede vestir de manera distinta durante 48 días sin repetir una combinación.

Vamos a considerar a continuación diferentes esquemas y contextos en donde es posible encontrar una fórmula matemática para ciertos problemas de conteo. En todos ellos aplicaremos el principio de multiplicación. El esquema general es el de extraer al azar k objetos, uno a la vez, de una urna con n objetos distintos. Esto se muestra en la Figura 1.8.

Ordenaciones con repetición: muestras con orden y con reemplazo

Suponga que tenemos una urna con n objetos distintos. Deseamos realizar k extracciones al azar de un objeto a la vez. Al efectuar una extracción, registramos el objeto

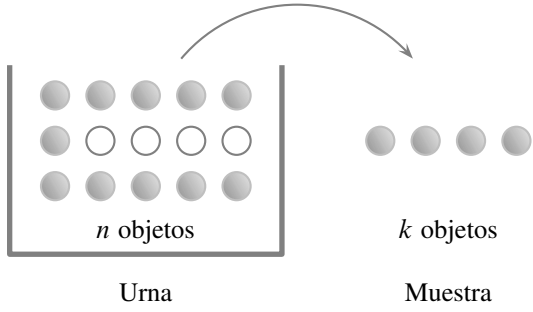


FIGURA 1.8

escogido y lo regresamos a la urna, de esta forma el mismo objeto puede ser extraído varias veces. El total de arreglos que se pueden obtener de esta urna al hacer k extracciones es el número n^k, pues en cada extracción tenemos n objetos posibles para escoger y efectuamos k extracciones. Esta fórmula es consecuencia del principio de multiplicación enunciado antes. A este número se le llama ordenaciones con repetición. Se dice que la muestra es con orden pues es importante el orden en el que se van obteniendo los objetos, y es con reemplazo pues cada objeto seleccionado se reincorpora a la urna.

EJEMPLO 1.20. Suponga que tenemos un conjunto de 60 caracteres diferentes. Este conjunto contiene todas las letras minúsculas del alfabeto, las letras mayúsculas, los diez dígitos y algunos caracteres especiales. ¿Cuántos passwords o palabras clave de longitud 4 se pueden construir usando este conjunto de 60 caracteres? Este es un ejemplo de una ordenación de 60 caracteres en donde se permiten las repeticiones. Como cada carácter de los 60 disponibles puede ser escogido para ser colocado en cada una de las cuatro posiciones de la palabra clave, entonces se pueden construir $60 \times 60 \times 60 \times 60 = 60^4 = 12,960,000$ distintos passwords de longitud 4.

Ordenaciones sin repetición: muestras con orden y sin reemplazo

Suponga que se tiene la misma situación que antes, una urna con n objetos y de los cuales se deben extraer, uno a uno, k objetos. Suponga esta vez que el muestreo es sin reemplazo, es decir, una vez seleccionado un objeto éste ya no se reincorpora a la urna. El total de arreglos distintos que se pueden obtener de este modo es el número: $n(n-1)(n-2)\cdots(n-k+1)$. Primeramente debemos observar que hay k factores en la expresión anterior. El primer factor es n y ello es debido a que tenemos cualesquiera de los n objetos para ser colocado en primera posición, para la segunda posición tenemos ahora $n-1$ objetos, para la tercera $n-2$ objetos, y así sucesivamente. Este razonamiento termina al escoger el k-ésimo objeto para el cual tenemos únicamente $n-k+1$ posibilidades. Nuevamente por el principio de multiplicación, la respuesta

es el producto indicado. La expresión encontrada puede escribirse como sigue:

$$P(n,k) = \frac{n!}{(n-k)!},$$

y se le llama permutaciones de n en k. En el caso particular cuando la muestra es exhaustiva, es decir, cuando $k = n$, o bien cuando todos los objetos son extraídos uno por uno, entonces se tienen todas las permutaciones o distintos órdenes en que se pueden colocar n objetos.

EJEMPLO 1.21. ¿De cuantas formas distintas pueden asignarse los premios primero, segundo y tercero en una rifa de 10 boletos numerados del 1 al 10? Claramente se trata de una ordenación sin repetición de 10 objetos en donde se deben extraer 3 de ellos. La respuesta es entonces que existen $10 \times 9 \times 8 = 720$ distintas asignaciones para los tres primeros lugares en la rifa.

Permutaciones: muestras exhaustivas con orden y sin reemplazo

La pregunta básica acerca del total de formas en que podemos poner en orden lineal (uno detrás de otro y por lo tanto no hay repetición) n objetos distintos tiene como respuesta el factorial de n, denotado por $n!$ y definido como sigue:

$$n! = n(n-1)(n-2)\cdots 3 \cdot 2 \cdot 1.$$

A este número también se le conoce como las permutaciones de n objetos, y se usa la notación $P(n) = n!$ Adicionalmente y por conveniencia se define $0! = 1$. Observe que las permutaciones de n objetos es un caso particular de la situación mencionada en la sección anterior sobre ordenaciones sin repetición cuando la muestra es exhaustiva, es decir, cuando se extraen uno a uno todos los objetos de la urna.

EJEMPLO 1.22. Si deseamos conocer el total de formas distintas en que podemos colocar una enciclopedia de 5 volúmenes en un librero, la respuesta es claramente $5! = 5 \times 4 \times 3 \times 2 \times 1 = 120$. El razonamiento es el siguiente: Cualquiera de los cinco libros puede ser colocado al principio, quedan cuatro libros por colocar en la segunda posición, restan entonces tres posibilidades para la tercera posición, etc. Por el principio de multiplicación la respuesta es el producto de estos números.

Combinaciones: muestras sin orden y sin reemplazo

Supongamos nuevamente que tenemos un conjunto de n objetos distinguibles y nos interesa obtener una muestra de tamaño k. Supongamos ahora que las muestras deben ser sin orden y sin reemplazo. Es decir, en la muestra no debe haber elementos repetidos, pues no hay reemplazo, y además la muestra debe verse como un conjunto pues no debe haber orden entre sus elementos. ¿Cuántas diferentes muestras podemos obtener de estas características? Para responder a esta pregunta seguiremos el siguiente

razonamiento: cuando el orden importa hemos encontrado antes la fórmula

$$\frac{n!}{(n-k)!}.$$

Ahora que no nos interesa el orden, observamos que cada uno de los arreglos de la fórmula anterior, está siendo contado $k!$ veces, las veces en que los mismos k elementos pueden ser permutados unos con otros, siendo que el conjunto de elementos es el mismo. Para obtener arreglos en donde el orden no importa, debemos entonces dividir por $k!$ La fórmula a la que hemos llegado se llama combinaciones de n en k y la denotaremos como sigue:

$$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}.$$

A este número también se le conoce con el nombre de coeficiente binomial de n en k, pues aparece en el famoso teorema del binomio: para cualesquiera números reales a y b, y para cualquier número entero $n \geq 0$,

$$(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k}\, b^k.$$

Para los casos $n = 2$ y $n = 3$ el teorema del binomio se reduce a las siguientes fórmulas que muy seguramente el lector conoce:

$$\begin{aligned} (a+b)^2 &= a^2 + 2ab + b^2. \\ (a+b)^3 &= a^3 + 3a^2b + 3ab^2 + b^3. \end{aligned}$$

EJEMPLO 1.23. ¿Cuántos equipos distintos de tres personas pueden escogerse de un grupo de 5 personas?
Repuesta: observe que el orden de las tres personas escogidas no es importante de modo que la respuesta es

$$\binom{5}{3} = \frac{5!}{3!\,(5-3)!} = 10.$$

El coeficiente binomial es también una forma de generar las entradas del así llamado triángulo de Pascal, que puede observarse en Figura 1.9. El n-ésimo renglón del triángulo de Pascal, iniciando desde cero, contiene los coeficientes del desarrollo de $(a+b)^n$. Existe una forma sencilla de construir este triángulo observando que cada uno de estos números, exceptuando los extremos, es la suma de los dos números inmediatos del renglón anterior. A este respecto véase por ejemplo el Ejercicio 69 en la página 148.

Coeficiente multinomial

Ahora consideremos que tenemos n objetos no necesariamente distintos unos de otros, por ejemplo, supongamos que tenemos k_1 objetos de un primer tipo, k_2 objetos de un segundo tipo, y así sucesivamente, hasta k_m objetos del tipo m, en donde

$k_1 + k_2 + \cdots + k_m = n$. Estos n objetos pueden todos ordenarse uno detrás de otro de tantas formas distintas como indica el así llamado coeficiente multinomial:

$$\binom{n}{k_1\, k_2\, \cdots\, k_{m-1}\, k_m} = \frac{n!}{k_1!\, k_2! \cdots k_{m-1}!\, k_m!}.$$

Un razonamiento para obtener esta fórmula es el siguiente: si consideramos que los

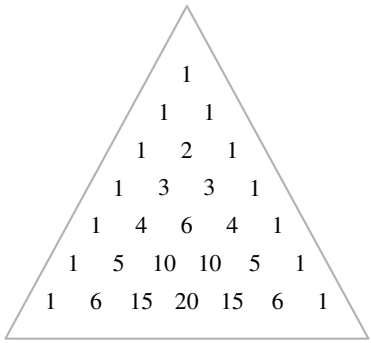


FIGURA 1.9. Triángulo de Pascal.

n objetos son todos distintos, entonces claramente las distintas formas en que pueden escribirse todos estos objetos uno detrás de otro es $n!$ Pero para cada uno de estos arreglos, los k_1 objetos del primer tipo, supuestos inicialmente distintos cuando en realidad no lo son, pueden permutarse entre sí de $k_1!$ formas diferentes, siendo que el arreglo total es el mismo. De aquí que debamos dividir por $k_1!$ Lo mismo sucede con los elementos del segundo tipo y así sucesivamente hasta los elementos del tipo m.

EJEMPLO 1.24. ¿Cuántas palabras distintas se pueden formar permutando las letras de la palabra "mamá" sin importar el acento?
Respuesta: Existen $\binom{4}{2\,2} = 6$ palabras distintas y éstas son:

mama	amma	mmaa
maam	amam	aamm.

El coeficiente multinomial aparece en la siguiente fórmula:

$$(a_1 + a_2 + \cdots + a_m)^n = \sum \binom{n}{k_1 \cdots k_m} a_1^{k_1} a_2^{k_2} \cdots a_m^{k_m}, \tag{1}$$

en donde la suma se efectúa sobre todos los posibles valores enteros no negativos de $k_1, k_2, \ldots, k_m$, tales que $k_1 + k_2 + \cdots + k_m = n$. A este resultado se le conoce como el teorema multinomial y es claramente una extensión del teorema del binomio. Por ejemplo, compruebe el lector que la fórmula (1) produce la siguiente expresión:

$$(a + b + c)^2 = a^2 + b^2 + c^2 + 2ab + 2ac + 2bc.$$

¿Puede usted desarrollar $(a+b+c)^3$? Es interesante observar que cuando hay únicamente dos tipos de objetos, el coeficiente multinomial se reduce al coeficiente binomial.

Muestras sin orden y con reemplazo

Finalmente consideremos el caso de hacer k extracciones de una urna de n objetos con las condiciones de que cada objeto extraído es regresado a la urna (y entonces puede ser elegido nuevamente), y en donde el orden de la muestra no es relevante. Para encontrar una fórmula para el total de muestras que pueden obtenerse con estas características usaremos una modelación distinta pero equivalente.

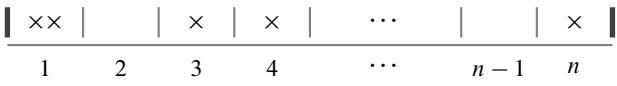


FIGURA 1.10

Consideremos el arreglo de n casillas de la Figura 1.10 junto con la siguiente interpretación: la primera casilla tiene dos cruces y eso indica que la bola uno fue seleccionada dos veces, la segunda casilla esta vacía y ello significa que la bola dos no fue seleccionada, etc. El número de cruces en la casilla i indica entonces el número de veces que la bola i fue seleccionada. En total debe haber k cruces pues es el total de extracciones. Deseamos entonces conocer el número de posibles arreglos que pueden obtenerse con estas características, y debe ser claro, después de algunos momentos de reflexión, que éste es el número de muestras de tamaño k, con reemplazo y sin orden, que se pueden obtener de un conjunto de n elementos distinguibles. Consideremos que las dos paredes en los extremos de este arreglo son fijas, estas paredes se encuentran ligeramente remarcadas como puede apreciarse en la Figura 1.10. Consideremos además que las posiciones intermedias, cruz o línea vertical, pueden moverse. En total hay $n+k-1$ objetos movibles y cambiar de posición estos objetos produce las distintas configuraciones posibles que nos interesan. El número total de estos arreglos es entonces

$$\binom{n+k-1}{k}$$

que equivale a colocar dentro de las $n+k-1$ posiciones las k cruces, dejando en los lugares restantes las paredes movibles.

Resumen de fórmulas

En el contexto de muestras de tamaño k tomadas de un conjunto de cardinalidad n, y a manera de resumen parcial, tenemos la tabla de fórmulas de la Figura 1.11.

Muestras	con reemplazo	sin reemplazo
con orden	n^k	$\frac{n!}{(n-k)!}$
sin orden	$\binom{n+k-1}{k}$	$\binom{n}{k}$

FIGURA 1.11

Debemos hacer énfasis, sin embargo, en que los problemas de conteo pueden ser difíciles de resolver y que para resolver un problema en particular no debemos clasificarlo forzosamente y de manera mecánica en alguno de los esquemas mencionados. Muy posiblemente el problema en cuestión requerirá de un razonamiento especial que involucre alguna combinación de las fórmulas encontradas. En algunos casos uno puede encontrar dos o mas "soluciones" distintas y aparentemente correctas de un problema. A veces las múltiples soluciones se deben a que el problema no esta bien especificado y por lo tanto pueden surgir ambigüedades en su interpretación. Véase el libro de Székely [21] para conocer una amplia gama de paradojas que surgen en la probabilidad y la estadística.

Otro aspecto que amerita mencionarse es que usaremos las técnicas de conteo principalmente para aplicarlas en la definición clásica de la probabilidad: $P(A) = \#A/\#\Omega$. Aunque estos métodos son bastante útiles, constituyen sólo una parte mínima de la teoría de la probabilidad. Mencionaremos también que la programación de computadoras puede ser una herramienta muy práctica para resolver problemas de conteo o simplemente para verificar la respuesta encontrada.

4. Probabilidad condicional e independencia

Los conceptos de probabilidad condicional e independencia surgieron de manera natural en el proceso de encontrar solución a algunos problemas provenientes de situaciones reales. En esta sección estudiaremos estos conceptos y demostraremos además dos resultados de amplia aplicación: el teorema de probabilidad total y el teorema de Bayes.

DEFINICIÓN 1.25. Sean A y B dos eventos en donde B es tal que su probabilidad es estrictamente positiva. La probabilidad condicional del evento A dado el evento B, denotada por $P(A \mid B)$, se define como sigue:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

La expresión $P(A \mid B)$ se lee probabilidad condicional del evento A dado el evento B, o simplemente probabilidad de A dado B. Es claro que para que la definición tenga sentido se necesita suponer que $P(B) > 0$. Por otro lado no existe definición para $P(A \mid B)$ cuando $P(B) = 0$. Ilustraremos a continuación el significado de la probabilidad condicional y comprobaremos que el evento B representa información adicional acerca del experimento aleatorio que modifica, en general, las probabilidades de los distintos eventos. El cálculo de probabilidades condicionadas a la ocurrencia del evento B tiene el efecto de reducir el espacio muestral Ω a únicamente el evento B, es decir, sólo los resultados contenidos en el evento B pueden ocurrir.

EJEMPLO 1.26. Considere el experimento de lanzar un dado equilibrado. Claramente el espacio muestral es $\Omega = \{1, 2, 3, 4, 5, 6\}$, el cual por hipótesis es equiprobable. Sean los eventos $A = \{2\}$ y $B = \{2, 4, 6\}$ = "Cae par". Entonces $P(A) = 1/6$ mientras que

$$P(A \mid B) = \frac{P(\{2\} \cap \{2, 4, 6\})}{P(\{2, 4, 6\})} = \frac{P(\{2\})}{P(\{2, 4, 6\})} = \frac{1/6}{3/6} = 1/3.$$

Observe que conocer la información de la ocurrencia del evento B, ha afectado la probabilidad del evento A, es decir, dada la información que el resultado del dado es un número par, la probabilidad de obtener "2" es ahora $1/3$.

DEFINICIÓN 1.27. Se dice que dos eventos cualesquiera A y B son independientes si se cumple la condición

$$P(A \cap B) = P(A)P(B).$$

Esta igualdad en la definición de independencia es equivalente a la expresión $P(A \mid B) = P(A)$ cuando $P(B) > 0$. La ventaja de esta última expresión es que posee una interpretación sencilla: dice que la probabilidad del evento A es la misma cuando sabemos que ha ocurrido el evento B (lado izquierdo), que cuando no sabemos nada (lado derecho). Es decir, la ocurrencia del evento B no afecta la probabilidad del evento A y por lo tanto son independientes. De manera análoga puede interpretarse la igualdad equivalente $P(B \mid A) = P(B)$, suponiendo naturalmente que $P(A) > 0$. En la mayoría de los casos de aplicación simplemente supondremos que dos eventos dados son independientes recurriendo únicamente a justificaciones intuitivas.

EJEMPLO 1.28. Suponga que se lanza una moneda dos veces. Es una hipótesis natural suponer que el resultado del primer lanzamiento no afecta el resultado del segundo lanzamiento. De este modo cualquier evento del primer ensayo es independiente de cualquier otro evento en el segundo ensayo.

EJEMPLO 1.29. Inicialmente uno podría asociar la idea de independencia de dos eventos con el hecho de que éstos sean ajenos, pero ello es erróneo en general. Considere por ejemplo A un evento cualquier y $B = \Omega$, entonces es claro que A y B son independientes pues $P(A \cap \Omega) = P(A)P(\Omega)$ y sin embargo, en general, $A \cap B \neq \emptyset$. Por lo tanto, el hecho de que dos eventos sean independientes no implica, en general, que sean ajenos.

EJEMPLO 1.30. Como complemento al ejemplo anterior, ilustraremos ahora el hecho de que dos eventos pueden ser ajenos sin ser independientes. Consideremos el experimento aleatorio de lanzar un dado y definamos los eventos A como obtener un número par y B como obtener un número impar. Es claro que los eventos A y B son ajenos, y sin embargo no son independientes pues $P(A \cap B) \neq P(A)P(B)$. Por lo tanto, dos eventos pueden ser ajenos y ello no implica necesariamente que sean independientes.

La definición de independencia de dos eventos puede generalizarse al caso de varios eventos de la siguiente forma.

DEFINICIÓN 1.31. Decimos que n eventos $A_1, A_2, \ldots, A_n$ son independientes si se satisfacen todas y cada una de las condiciones siguientes:

$$\begin{aligned} P(A_i \cap A_j) &= P(A_i)P(A_j), \quad i, j \text{ distintos.} \\ P(A_i \cap A_j \cap A_k) &= P(A_i)P(A_j)P(A_k), \quad i, j, k \text{ distintos.} \\ &\vdots \\ P(A_1 \cap \cdots \cap A_n) &= P(A_1) \cdots P(A_n). \end{aligned}$$

En general, para verificar que n eventos son independientes es necesario comprobar todas y cada una de las igualdades arriba enunciadas. Es decir, cualquiera de estas igualdades no implica, en general, la validez de alguna otra, es necesario pues verificarlas todas. No es difícil darse cuenta que el total de igualdades es $2^n - n - 1$. ¿Puede usted justificar este resultado?

EJEMPLO 1.32. Sea $\Omega = \{1, 2, 3, 4\}$ un espacio muestral equiprobable. Sean los eventos $A = \{1, 2\}$, $B = \{2, 3\}$ y $C = \{2, 4\}$. ¿Son A, B y C independientes? No lo son, pues aunque se cumplen las igualdades $P(A \cap B) = P(A)P(B)$, $P(A \cap C) = P(A)P(C)$, y $P(B \cap C) = P(B)P(C)$, no se cumple que $P(A \cap B \cap C) = P(A)P(B)P(C)$.

Teorema de probabilidad total

Antes de enunciar el siguiente resultado recordaremos el concepto de partición de un conjunto: Una partición finita de un conjunto Ω es una colección $B_1, \ldots, B_n$ de subconjuntos de Ω tal que cada uno de estos conjuntos es distinto del vacío, la colección es disjunta dos a dos, esto es, para índices i y j distintos, se cumple que $B_i \cap B_j \neq 0$, y además la unión de toda la colección produce el total Ω, es decir, $B_1 \cup B_2 \cup \cdots \cup B_n = \Omega$. En la Figura 1.12 se muestra gráficamente el concepto de partición de un conjunto. Ahora podemos enunciar y demostrar el muy útil teorema de probabilidad total.

TEOREMA 1.33 (Teorema de probabilidad total). *Sea $B_1, B_2, \ldots, B_n$ una partición de Ω tal que cada elemento de la partición tiene probabilidad estrictamente positiva.*

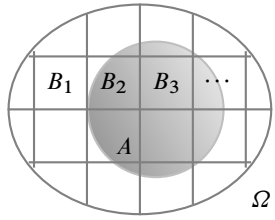


FIGURA 1.12

Para cualquier evento A,

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i) P(B_i).$$

DEMOSTRACIÓN. Primero observemos que el evento A puede escribirse como sigue

$$A = A \cap \Omega = A \cap \bigcup_{i=1}^{n} B_i = \bigcup_{i=1}^{n} (A \cap B_i),$$

en donde los eventos $A \cap B_1, \ldots, A \cap B_n$ son ajenos dos a dos, de modo que

$$P(A) = P(\bigcup_{i=1}^{n} A \cap B_i) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A \mid B_i) P(B_i). \qquad \blacksquare$$

Observe que cuando la partición de Ω consta de únicamente dos elementos, B y B^c, la fórmula del teorema de probabilidad total se reduce a la siguiente expresión:

$$P(A) = P(A|B)\, P(B) + P(A|B^c)\, P(B^c).$$

EJEMPLO 1.34. Suponga que tenemos dos cajas: una con 3 bolas blancas y 7 bolas de color gris, la otra con 6 blancas y 6 grises. Esta situación se ilustra en la Figura 1.13. Si se elije una caja al azar y después se saca una bola, ¿cuál es la probabilidad de que sea blanca?

FIGURA 1.13

El experimento aleatorio consiste entonces en escoger una caja al azar, con idéntica probabilidad cada una de ellas, y después escoger una bola de la caja escogida. Es claro entonces que el espacio muestral puede escribirse como sigue

$$\Omega = \{(C_1, B), (C_1, G), (C_2, B), (C_2, G)\},$$

en donde C_1 y C_2 denotan los eventos en donde las cajas uno y dos fueron escogidas, respectivamente, y B y G denotan los eventos en donde una bola blanca o gris fueron escogidas respectivamente. Nos piden calcular la probabilidad de B. Observe que es fácil calcular la probabilidad de este evento cuando se conoce la caja que fue escogida. Esto sugiere condicionar sobre el resultado de escoger alguna de las dos cajas y aplicar el teorema de probabilidad total, es decir,

$$\begin{aligned} P(B) &= P(B \mid C_1)P(C_1) + P(B \mid C_2)P(C_2) \\ &= \frac{3}{10} \times \frac{1}{2} + \frac{6}{12} \times \frac{1}{2} \\ &= \frac{2}{5}. \end{aligned}$$

Observe además que la partición del espacio muestral consta de dos elementos:

$$\{(C_1, B), (C_1, G)\} \quad \text{y} \quad \{(C_2, B), (C_2, G)\}.$$

¿Puede usted comprobar que $P(G) = 3/5$? Puede uno también preguntarse por situaciones aparentemente extrañas como la siguiente: si se obtuvo una bola blanca, ¿cuál es la probabilidad de que haya sido obtenida de la primera caja? Es posible calcular esta probabilidad a través del teorema de Bayes.

EJEMPLO 1.35. Suponga que en una población humana de igual número de hombres y mujeres, el 4 % de hombres son daltónicos y el 1 % de las mujeres son daltónicas. Una persona es elegida al azar, ¿Cuál es la probabilidad de que sea daltónica? Definamos primero los eventos de interés. Sea M el evento "La persona escogida es mujer", H el evento "La persona escogida es hombre" y D el evento "La persona escogida es daltónica". Deseamos calcular $P(D)$. Por el teorema de probabilidad total,

$$\begin{aligned} P(D) &= P(D \mid M)P(M) + P(D \mid H)P(H) \\ &= \frac{1}{100} \times \frac{1}{2} + \frac{4}{100} \times \frac{1}{2} \\ &= \frac{1}{40}. \end{aligned}$$

Teorema de Bayes

El resultado interesante que estudiaremos a continuación involucra nuevamente probabilidades condicionales. Fue publicado por primera vez en 1763, dos años después de la muerte de su creador: el matemático y teólogo inglés Thomas Bayes.

Thomas Bayes
(Inglaterra 1702–1761)

TEOREMA 1.36 (Teorema de Bayes). *Sea $B_1, \ldots, B_n$ una partición de Ω tal que cada elemento de la partición tiene probabilidad estrictamente positiva. Sea A un evento tal que $P(A) > 0$. Entonces para cada $j = 1, 2, \ldots, n$,*

$$P(B_j \mid A) = \frac{P(A \mid B_j)P(B_j)}{\sum_{i=1}^{n} P(A \mid B_i)P(B_i)}.$$

DEMOSTRACIÓN. Por la definición de probabilidad condicional y el teorema de probabilidad total tenemos que

$$P(B_j \mid A) = \frac{P(A \cap B_j)}{P(A)} = \frac{P(A \mid B_j)P(B_j)}{P(A)} = \frac{P(A \mid B_j)P(B_j)}{\sum_{i=1}^{n} P(A \mid B_i)P(B_i)}. \quad \blacksquare$$

En la Figura 1.12 se encuentra una representación gráfica de una partición del espacio muestral y un evento A. Nuevamente observamos que en el caso cuando la partición de Ω consta de sólo dos elementos: B y B^c, el teorema de Bayes, para el evento B, adquiere la forma:

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A \mid B)P(B) + P(A \mid B^c)P(B^c)}.$$

EJEMPLO 1.37. En una fábrica hay dos máquinas. La máquina 1 realiza el 60 % de la producción total y la máquina 2 el 40 %. De su producción total, la máquina 1 produce 3 % de material defectuoso, la 2 el 5 %. El asunto es que se ha encontrado un material defectuoso, ¿cuál es la probabilidad de que este material defectuoso provenga de la máquina 2?

Solución: Sea M_1 el evento "La máquina 1 produjo el material escogido", M_2 en evento "La máquina 2 produjo el material escogido" y finalmente sea D el evento "El material escogido es defectuoso". El problema es encontrar $P(M_2 \mid D)$ y observamos que la

información que tenemos es $P(D \mid M_2)$. Por el teorema de Bayes tenemos entonces que

$$\begin{aligned} P(M_2 \mid D) &= \frac{P(D \mid M_2)P(M_2)}{P(D \mid M_1)P(M_1) + P(D \mid M_2)P(M_2)} \\ &= \frac{\frac{5}{100} \times \frac{40}{100}}{\frac{3}{100} \times \frac{60}{100} + \frac{5}{100} \times \frac{40}{100}} \\ &= \frac{10}{19}. \end{aligned}$$

Como un ejercicio al lector se deja comprobar que $P(M_1 \mid D) = 9/19$.

EJEMPLO 1.38. En un laboratorio se descubrió una prueba para detectar cierta enfermedad y sobre la eficacia de dicha prueba se conoce lo siguiente: si se denota por E el evento de que un paciente tenga la enfermedad y por N el evento de que la prueba resulte negativa, entonces se sabe que $P(N^c \mid E) = 0.95$, $P(N \mid E^c) = 0.96$ y $P(E) = 0.01$. Con esta información uno podría pensar que la prueba es muy buena, sin embargo calcularemos las probabilidades $P(E \mid N)$ y $P(E \mid N^c)$ usando el teorema de Bayes.

$$\begin{aligned} P(E \mid N) &= \frac{P(N \mid E)P(E)}{P(N \mid E)P(E) + P(N \mid E^c)P(E^c)} \\ &= \frac{0.05 \times 0.01}{0.05 \times 0.01 + 0.96 \times 0.99} \\ &= 0.000526\,. \end{aligned}$$

Es bueno que esta probabilidad sea pequeña, pero por otro lado,

$$\begin{aligned} P(E \mid N^c) &= \frac{P(N^c \mid E)P(E)}{P(N^c \mid E)P(E) + P(N^c \mid E^c)P(E^c)} \\ &= \frac{0.95 \times 0.01}{0.95 \times 0.01 + 0.04 \times 0.99} \\ &= 0.193\,. \end{aligned}$$

Esta última probabilidad es demasiado pequeña y por lo tanto la prueba no es muy confiable en tales casos.

5. Variables aleatorias

Dado un experimento aleatorio cualquiera, una variable aleatoria es una transformación X del espacio de resultados Ω al conjunto de números reales, esto es,

$$X : \Omega \to \mathbb{R}.$$

A menudo se escribe simplemente v.a. en lugar del término variable aleatoria. En sentido estricto una variable aleatoria es una función de Ω en $\mathbb{R}$ que satisface además cierta condición de medibilidad, pero omitiremos tales tecnicismos pues no son de utilidad para nuestros propósitos en este texto. Suponga entonces que se efectúa el experimento aleatorio una vez y se obtiene un resultado ω en Ω. Al transformar este resultado con la variable aleatoria X se obtiene un número real $X(\omega) = x$. Podemos entonces suponer que los posibles resultados del experimento aleatorio son los diferentes números reales x que la función X puede tomar. Ilustramos de manera gráfica el concepto de variable aleatoria en la Figura1.14. Debemos hacer aquí la siguiente observación importante: seguiremos la notación usual de utilizar la letra mayúscula X para denotar una variable aleatoria cualquiera, es decir, X es una función de Ω en $\mathbb{R}$, mientras que la letra minúscula x denota un número real y representa un posible valor de la variable aleatoria. En general, las variables aleatorias se denotan usando las últimas letras del alfabeto en mayúsculas, U, V, W, X, Y, Z, y para un valor cualquiera de ellas se usa la misma letra pero en minúscula.

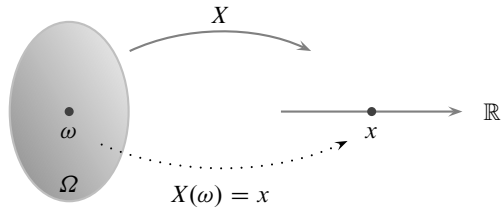


FIGURA 1.14

EJEMPLO 1.39. Suponga que un experimento aleatorio consiste en lanzar al aire una moneda y observar la cara superior una vez que la moneda cae. Denotemos por "Cara" y "Cruz" los dos lados de la moneda. Entonces claramente el espacio muestral es el conjunto $\Omega = \{\text{"Cara"}, \text{"Cruz"}\}$. Defina la variable aleatoria $X : \Omega \to \mathbb{R}$ de la siguiente forma

$$\begin{aligned} X(\text{"Cara"}) &= 0, \\ X(\text{"Cruz"}) &= 1. \end{aligned}$$

De este modo podemos suponer que el experimento aleatorio tiene dos valores numéricos posibles: 0 y 1. Observe que los números 0 y 1 son en realidad arbitrarios, otro par de números distintos puede ser escogido para distinguir los dos resultados del

experimento aleatorio. Podemos también definir la variable aleatoria $Y : \Omega \to \mathbb{R}$ de la siguiente forma

$$\begin{aligned} Y(\text{“Cara”}) &= 2, \\ Y(\text{“Cruz”}) &= 2. \end{aligned}$$

En este caso la variable Y solo toma un valor: el número 2. Cualquier resultado del experimento aleatorio produce, a través de la función Y, el número 2. Decimos que Y es la variable aleatoria constante 2.

EJEMPLO 1.40. Consideremos el experimento aleatorio consistente en lanzar un dardo en un tablero circular de radio uno como se muestra en la Figura 1.15. El espacio muestral o conjunto de posibles resultados del experimento se puede escribir como sigue: $\Omega = \{(x, y) : x^2 + y^2 \leq 1\}$.

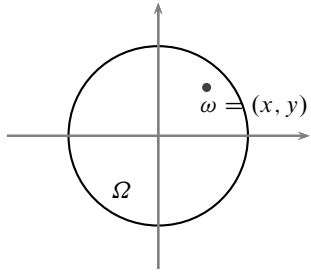


FIGURA 1.15

Los siguientes son ejemplos de variables aleatorias, es decir, funciones de Ω en $\mathbb{R}$, asociadas a este experimento aleatorio.

a) $X(x, y) = x$, proyección sobre el eje horizontal.
b) $Y(x, y) = y$, proyección sobre el eje vertical.
c) $Z(x, y) = \sqrt{x^2 + y^2}$, distancia al centro del círculo.
d) $V(x, y) = |x| + |y|$, distancia del taxista.
e) $W(x, y) = xy$, producto de las coordenadas.

Variables discretas vs variables continuas

Considerando el conjunto de valores que una variable aleatoria puede tomar, vamos a clasificar a las variables aleatorias en dos tipos: discretas o continuas. Decimos que una v.a. es discreta cuando el conjunto de valores que ésta toma es un conjunto discreto, es decir, un conjunto finito o numerable. Por ejemplo, el conjunto $\{0, 1, 2, \ldots, n\}$ es un conjunto discreto porque es finito, lo mismo $\mathbb{N}$ pues aunque es infinito, es numerable y por lo tanto discreto. Por otra parte, decimos que una variable aleatoria es continua cuando toma todos los valores dentro de un intervalo $(a, b) \subseteq \mathbb{R}$. Esta clasificación de

variables aleatorias no es completa pues existen variables que no son de ninguno de los dos tipos mencionados. Sin embargo, por simplicidad, en este curso estudiaremos únicamente variables aleatorias que son discretas o continuas.

EJEMPLO 1.41. Las variables X y Y definidas líneas arriba en el ejemplo del lanzamiento de una moneda, son variables aleatorias discretas. En ese ejemplo el espacio muestral mismo es discreto y por lo tanto las variables aleatorias que pueden allí definirse tienen que ser discretas forzosamente. En el ejemplo del lanzamiento de un dardo en un tablero circular de radio uno, el espacio muestral (Figura 1.15) es infinito no numerable, las variables X, Y, Z, V y W definidas allí son todas variables aleatorias continuas. Si se dibujan círculos concéntricos alrededor del origen y si se asignan premios asociados a cada una de las regiones resultantes, puede obtenerse un ejemplo de una variable aleatoria discreta sobre este espacio muestral.

EJEMPLO 1.42. Un experimento aleatorio consiste en escoger a una persona ω al azar. La variable aleatoria X evaluada en ω corresponde a conocer la siguiente característica o una codificación de esta característica de la persona escogida. En cada caso se trata de una variable aleatoria discreta:

a) Edad en años.
b) Número de hijos.
c) Peso.
d) Estatura.
e) Sueldo.

Eventos y variables aleatorias

En lo sucesivo emplearemos la siguiente notación importante: si A es un subconjunto de $\mathbb{R}$, entonces la expresión $(X \in A)$, incluyendo el paréntesis, denota el conjunto $\{\omega \in \Omega : X(\omega) \in A\}$, es decir,

$$(X \in A) = \{\omega \in \Omega : X(\omega) \in A\}.$$

En palabras, la expresión $(X \in A)$ denota aquel conjunto de elementos ω del espacio muestral Ω tales que bajo la aplicación de la función X toman un valor dentro del conjunto A. A este conjunto se le llama la imagen inversa de A y se le denota por $X^{-1}A$, lo cual no debe confundirse con la función inversa. Por ejemplo, consideremos que el conjunto A es el intervalo (a, b), entonces el evento $(X \in (a, b))$ también puede escribirse como $(a < X < b)$ y ambos son una abreviación del evento

$$\{\, \omega \in \Omega : a < X(\omega) < b \,\}.$$

Como otro ejemplo que será de nuestro interés más adelante, considere que el conjunto A es el intervalo infinito $(-\infty, x]$ para alguna x en $\mathbb{R}$. Entonces el evento $(X \in (-\infty, x])$ también puede escribirse como $(X \leq x)$ y significan

$$\{\, \omega \in \Omega : -\infty < X(\omega) \leq x \,\}.$$

El siguiente ejemplo muestra casos concretos de uso de esta notación para denotar eventos junto con sus respectivas probabilidades.

EJEMPLO 1.43. Consideremos nuevamente el experimento de lanzar una moneda y la variable aleatoria X que lleva el resultado "Cara" al valor 0 y el resultado "Cruz" al valor 1. Tenemos por ejemplo que $(X \in [1,\infty)) = \{\text{"Cruz"}\}$ pues el conjunto de elementos de Ω tales que bajo la función X toman un valor mayor o igual a uno, es decir, caen dentro del intervalo $[1,\infty)$, es únicamente el elemento "Cruz". Por lo tanto $P(X \in [1,\infty)) = P\{\text{"Cruz"}\} = 1/2$. Del mismo modo puede verificarse que

a) $P(X \in [1,2)) = P(\{\text{"Cruz"}\}) = 1/2$.
b) $P(X \in [0,1)) = P(\{\text{"Cara"}\}) = 1/2$.
c) $P(X \in [2,4]) = P(\emptyset) = 0$.
d) $P(X = 1) = P(\{\text{"Cruz"}\}) = 1/2$.
e) $P(X \leq -1) = P(\emptyset) = 0$.
f) $P(X \geq 0) = P(\Omega) = 1$.

En lo que resta del curso se usará con mucha frecuencia la notación arriba explicada. El lector debe asegurarse de comprender bien que si x es un número real entonces $(X \leq x)$ es un subconjunto de Ω y por lo tanto un evento. Lo mismo sucede con el complemento de este conjunto que es $(X > x)$. Podemos escribir entonces la igualdad de conjuntos $(X \leq x) \cup (X > x) = \Omega$. Y aplicando probabilidad se obtiene $P(X \leq x) + P(X > x) = 1$.

Nota importante. A través de una variable aleatoria se puede considerar que los posibles resultados de un experimento aleatorio no son elementos ω en Ω sino números reales que la variable aleatoria puede tomar. Esta es una consideración radical pues ya no consideraremos experimentos aleatorios particulares, ni espacios muestrales arbitrarios Ω, ni eventos (subconjuntos) de Ω; en lugar de ello consideraremos que una cierta variable aleatoria de interés toma valores en un cierto subconjunto de números reales. La probabilidad definida antes para subconjuntos de Ω se traslada, como explicamos antes, a probabilidades para subconjuntos de $\mathbb{R}$. Esta perspectiva permite estudiar modelos generales y después aplicarlos a situaciones particulares. A partir de ahora, el término variable aleatoria aparecerá con bastante frecuencia en el presente texto.

6. Funciones de densidad y de distribución

En esta sección vamos a explicar la forma de asociar a cada variable aleatoria dos funciones que nos proveen de información acerca de las características de la variable aleatoria. Estas funciones, llamadas función de densidad y función de distribución, nos permiten representar a un mismo tiempo tanto los valores que puede tomar la variable como las probabilidades de los distintos eventos. Definiremos primero la función de densidad para una variable aleatoria discreta, después para una continua, y finalmente definiremos la función de distribución para ambos tipos de variables aleatorias.

Función de probabilidad para una v.a. discreta

Sea X una variable aleatoria discreta que toma los valores $x_0, x_1, \ldots$ con probabilidades respectivas $P(X = x_0), P(X = x_1), \ldots$. Esta lista de valores numéricos y sus probabilidades puede ser finita o bien infinita, pero numerable. La función de probabilidad de la variable X denotada por $f(x) : \mathbb{R} \to [0, \infty)$ se define como sigue

$$f(x) = \begin{cases} P(X = x) & \text{si } x = x_0, x_1, \ldots \\ 0 & \text{otro caso.} \end{cases} \tag{2}$$

En palabras, la función de probabilidad es simplemente aquella función que indica la probabilidad en los distintos valores que toma la variable aleatoria. Recordemos que es importante poder distinguir entre X y x, pues conceptualmente son cosas muy distintas. Denotaremos generalmente a una función de probabilidad con la letra f minúscula. A veces escribiremos $f_X(x)$ y el subíndice nos ayudará a especificar que tal función es la función de probabilidad de la variable X. Esta notación será particularmente útil cuando consideremos varias variables aleatorias a la vez. A toda función de la forma (2) la llamaremos función de probabilidad. Observe que se cumplen las siguientes dos propiedades:

a) $f(x) \geq 0$, para toda $x \in \mathbb{R}$.

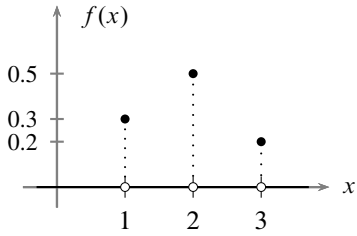
b) $\sum_{i=0}^{\infty} f(x_i) = 1$.

Recíprocamente, a toda función $f(x) : \mathbb{R} \to [0, 1]$ que sea cero excepto en ciertos puntos $x_0, x_1, \ldots$ en donde la función toma valores positivos se le llamará función de probabilidad si cumple con las dos propiedades anteriores.

EJEMPLO 1.44. Considere la variable aleatoria discreta X que toma los valores 1, 2 y 3, con probabilidades 0.3, 0.5 y 0.2 respectivamente. Observe que no se especifica ni el experimento aleatorio ni el espacio muestral, únicamente los valores de la variable aleatoria y las probabilidades asociadas. La función de probabilidad de X es

$$f(x) = \begin{cases} 0.3 & \text{si } x = 1, \\ 0.5 & \text{si } x = 2, \\ 0.2 & \text{si } x = 3, \\ 0 & \text{otro caso.} \end{cases}$$

Esta función se muestra gráficamente en la Figura 1.16 (a). Alternativamente podemos también expresar esta función mediante la tabla de la Figura 1.16 (b). En esta representación se entiende de manera implícita que $f(x)$ es cero para cualquier valor de x distinto de 1, 2 y 3. En particular, no debe ser difícil para el lector comprobar que las siguientes probabilidades son correctas: $P(X \geq 2) = 0.7$, $P(|X| = 1) = 0.3$ y $P(X < 1) = 0$.



(a)

x	1	2	3
$f(x)$	0.3	0.5	0.2

(b)

FIGURA 1.16

EJEMPLO 1.45. Encontraremos el valor de la constante c que hace que la siguiente función sea de probabilidad.

$$f(x) = \begin{cases} c\,x & \text{si } x = 0, 1, 2, 3, \\ 0 & \text{otro caso.} \end{cases}$$

Los posibles valores de la variable aleatoria discreta (no especificada) son $0, 1, 2$ y 3, con probabilidades $0, c, 2c$ y $3c$, respectivamente. Como la suma de estas probabilidades debe ser uno, obtenemos la ecuación $c + 2c + 3c = 1$. De aquí obtenemos $c = 1/6$. Este es el valor de c que hace que $f(x)$ sea no negativa y sume uno, es decir, una función de probabilidad.

Función de densidad para una v.a. continua

Sea X una variable aleatoria continua. Decimos que la función integrable y no negativa $f(x) : \mathbb{R} \to [0, \infty)$ es la función de densidad de X si para cualquier intervalo $[a, b]$ de $\mathbb{R}$ se cumple la igualdad

$$P(X \in [a, b]) = \int_a^b f(x)\,dx.$$

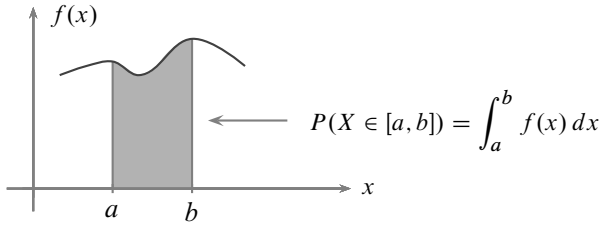


FIGURA 1.17. La probabilidad como un área.

Es decir, la probabilidad de que la variable tome un valor dentro del intervalo $[a, b]$ se puede calcular o expresar como el área bajo la función de densidad en dicho intervalo. De esta forma el cálculo de una probabilidad se reduce al cálculo de una integral. Véase la Figura 1.17 en donde se muestra esta forma de calcular probabilidades. No es difícil comprobar que toda función de densidad $f(x)$ de una variable aleatoria continua cumple las siguientes propiedades análogas al caso discreto.

a) $f(x) \geq 0$, para toda $x \in \mathbb{R}$.

b) $\displaystyle\int_{-\infty}^{\infty} f(x)\,dx = 1$.

Recíprocamente, toda función $f(x) : \mathbb{R} \to [0, \infty)$ que satisfaga estas dos propiedades (sin necesidad de tener una variable aleatoria de por medio) se llamará función de densidad.

EJEMPLO 1.46. La función dada por

$$f(x) = \begin{cases} 1/2 & \text{si } x \in (1, 3), \\ 0 & \text{otro caso,} \end{cases}$$

es la función de densidad de una variable aleatoria continua que toma valores en el intervalo $(1, 3)$ y cuya gráfica aparece en la Figura 1.18. Observe que se trata de una función no negativa y cuya integral vale uno.

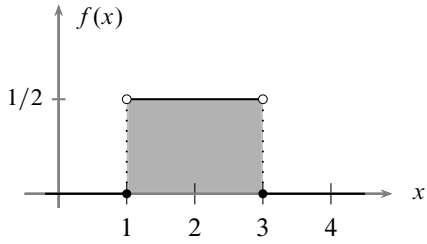


FIGURA 1.18

EJEMPLO 1.47. Encontraremos el valor de la constante c que hace que la siguiente función sea de densidad.

$$f(x) = \begin{cases} c\,|x| & \text{si } x \in [-1, 1], \\ 0 & \text{otro caso.} \end{cases}$$

Se trata de una variable aleatoria continua que toma valores en el intervalo $[-1, 1]$. Como esta función debe integrar uno tenemos que

$$1 = \int_{-1}^{1} c\,|x|\,dx = c \int_{0}^{1} 2x\,dx = c.$$

Por lo tanto, cuando tomamos $c = 1$ esta función resulta ser una función de densidad pues ahora cumple con ser no negativa e integrar uno.

Función de distribución

Sea X una variable aleatoria discreta o continua. La función de distribución de X, denotada por $F(x) : \mathbb{R} \to [0, 1]$, se define como

$$F(x) = P(X \leq x). \tag{3}$$

Esto es, la función de distribución evaluada en un número x cualquiera es simplemente la probabilidad de que la variable aleatoria tome un valor menor o igual a x, o en otras palabras, que tome un valor en el intervalo $(-\infty, x]$. Siendo $F(x)$ una probabilidad, sus valores están siempre entre cero y uno. En el caso discreto, suponiendo que $f(x)$ es la función de probabilidad de X, la función de distribución (3) se calcula como sigue

$$F(x) = \sum_{u \leq x} f(u).$$

En el caso continuo, si $f(x)$ es la función de densidad de X, por (3) se tiene que

$$F(x) = \int_{-\infty}^{x} f(u)\, du.$$

Esta función resulta ser importante y se le conoce también, por razones evidentes, con el nombre de función de acumulación de probabilidad. Con un par de ejemplo mostraremos la forma de calcular esta función a partir de la función de probabilidad o de densidad.

EJEMPLO 1.48 (Caso discreto). Considere la variable aleatoria discreta X de la Figura 1.16. Tenemos que la correspondiente función de distribución evaluada en x se calcula sumando las probabilidades $P(X = u)$ para valores de u menores o iguales a x, es decir,

$$F(x) = P(X \leq x) = \sum_{u \leq x} P(X = u) = \begin{cases} 0 & \text{si } x < 1, \\ 0.3 & \text{si } 1 \leq x < 2, \\ 0.8 & \text{si } 2 \leq x < 3, \\ 1 & \text{si } x \geq 3, \end{cases}$$

cuya gráfica aparece en la Figura 1.19 (a). Este es el comportamiento típico de una función de distribución discreta, es no decreciente, constante por pedazos, y si la función tiene una discontinuidad en x, entonces el tamaño de tal discontinuidad es exactamente la probabilidad de que la variable aleatoria tome ese valor.

EJEMPLO 1.49 (Caso continuo). Considere ahora la variable aleatoria continua X de la Figura 1.18. La correspondiente función de distribución se obtiene calculando la siguiente integral separando el cálculo en tres casos distintos:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\, du = \begin{cases} 0 & \text{si } x \leq 1, \\ (x-1)/2 & \text{si } 1 < x < 3. \\ 1 & \text{si } x \geq 3, \end{cases}$$

La gráfica de esta función aparece en la Figura 1.19 (b). Observe que esta función es continua y no decreciente.

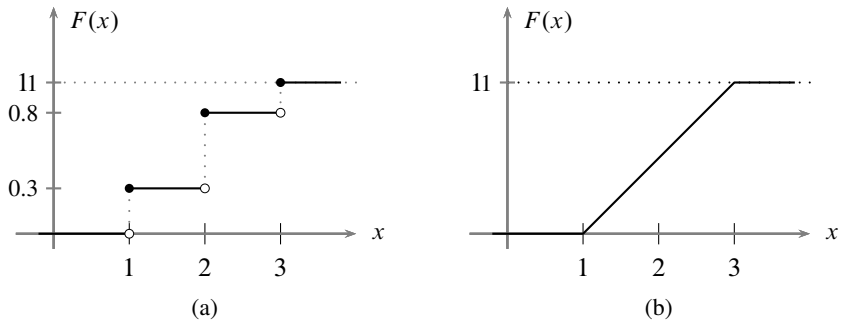


FIGURA 1.19

EJEMPLO 1.50 (Caso discreto). Considere la siguiente función de probabilidad

$$f(x) = \begin{cases} 1/6 & \text{si } x = 1, 2, 3, 4, 5, 6, \\ 0 & \text{otro caso.} \end{cases}$$

Considerando los distintos valores para x, puede encontrarse la función de distribución $F(x)$, la cual está dada por

$$F(x) = P(X \leq x) = \sum_{u \leq x} f(u) = \begin{cases} 0 & \text{si } x < 1, \\ 1/6 & \text{si } 1 \leq x < 2, \\ 2/6 & \text{si } 2 \leq x < 3, \\ 3/6 & \text{si } 3 \leq x < 4, \\ 4/6 & \text{si } 4 \leq x < 5, \\ 5/6 & \text{si } 5 \leq x < 6, \\ 1 & \text{si } x \geq 6, \end{cases}$$

Se deja como ejercicio al lector graficar con precisión las funciones $f(x)$ y $F(x)$.

EJEMPLO 1.51 (Caso continuo). Considere la siguiente función de densidad

$$f(x) = \begin{cases} |x| & \text{si } x \in [-1, 1], \\ 0 & \text{otro caso.} \end{cases}$$

Integrando esta función desde menos infinito hasta x, para distintos valores de x, se encuentra la función de distribución

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\, du = \begin{cases} 0 & \text{si } x < -1, \\ (1 - x^2)/2 & \text{si } -1 \leq x < 0, \\ (1 + x^2)/2 & \text{si } 0 \leq x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$

Nuevamente se deja como ejercicio al lector la graficación de $f(x)$ y $F(x)$ para un mejor entendimiento de la relación que guardan estas dos funciones.

En los ejemplos anteriores se ha mostrado la forma de obtener $F(x)$ a partir de $f(x)$. Ahora explicaremos el proceso contrario, es decir, obtener $f(x)$ a partir de $F(x)$. En el caso continuo tenemos que para toda x en $\mathbb{R}$,

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(u)\, du,$$

de modo que por el Teorema Fundamental del Cálculo, y cuando $F(x)$ es diferenciable, tenemos que

$$F'(x) = f(x).$$

De este modo podemos encontrar $f(x)$ a partir de $F(x)$. En el siguiente ejemplo se muestra la forma de llevar a cabo este procedimiento.

EJEMPLO 1.52. Considere la función de distribución

$$F(x) = \begin{cases} 0 & \text{si } x < -1, \\ (1+x^3)/2 & \text{si } -1 \leq x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$

Observamos que se trata de una función continua y diferenciable. Derivando entonces en cada una de las tres regiones de definición se encuentra que

$$f(x) = \begin{cases} 3x^2/2 & \text{si } -1 < x < 1, \\ 0 & \text{otro caso.} \end{cases}$$

En el caso discreto, la función de probabilidad se obtiene de la función de distribución del siguiente modo:

$$f(x) = P(X = x) = F(x) - F(x-),$$

en donde $F(x-)$ es el límite por la izquierda de la función F en el punto x. Análogamente, la expresión $F(x+)$ significa el límite por la derecha de la función F en el punto x. En símbolos,

$$\begin{aligned} F(x-) &= \lim_{h \to 0} F(x-h), \quad \text{con } h > 0, \\ F(x+) &= \lim_{h \to 0} F(x+h), \quad \text{con } h > 0. \end{aligned}$$

EJEMPLO 1.53. Considere la función de distribución

$$F(x) = \begin{cases} 0 & \text{si } x < -1, \\ 1/3 & \text{si } -1 \leq x < 0, \\ 2/3 & \text{si } 0 \leq x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$

Al graficar esta función uno puede darse cuenta que se trata de una función constante por pedazos. Los puntos donde esta función tiene incrementos y los tamaños de estos incrementos determinan la correspondiente función de probabilidad, la cual está dada por

$$f(x) = \begin{cases} 1/3 & \text{si } x = -1, 0, 1, \\ 0 & \text{otro caso.} \end{cases}$$

Mencionaremos a continuación algunas propiedades generales válidas para toda función de distribución, sea ésta discreta o continua.

PROPOSICIÓN 1.54. *Toda función de distribución $F(x)$ satisface las siguientes propiedades:*

a) $\lim_{x\to\infty} F(x) = 1$.
b) $\lim_{x\to-\infty} F(x) = 0$.
c) *Si* $x_1 \leq x_2$, *entonces* $F(x_1) \leq F(x_2)$.
d) $F(x) = F(x+)$.

DEMOSTRACIÓN.

a) Cuando x tiende a infinito el conjunto $(X \leq x)$ se aproxima al conjunto $(X \leq \infty)$ que es idéntico a Ω, por lo tanto, cuando $x \to \infty$, $F(x) \to P(X \leq \infty) = P(\Omega) = 1$.
b) Análogamente el conjunto $(X \leq x)$ se aproxima al conjunto $(X \leq -\infty) = \emptyset$ cuando x tiende a menos infinito. Por lo tanto, cuando $x \to -\infty$, $F(x) \to P(X \leq -\infty) = P(\emptyset) = 0$.
c) Observe que el evento $(x_1 < X \leq x_2)$ puede descomponerse en la diferencia $(X \leq x_2) - (X \leq x_1)$, en donde $(X \leq x_1) \subseteq (X \leq x_2)$. Por lo tanto $P(x_1 < X \leq x_2) = P(X \leq x_2) - P(X \leq x_1) = F(x_2) - F(x_1)$. Como la probabilidad es no negativa, se concluye que $F(x_1) \leq F(x_2)$.
d) Para $h > 0$ tenemos que $F(x+h) = P(X \leq x+h) = P(X \leq x) + P(x < X \leq x+h)$, de modo que cuando h tiende a cero, el conjunto $(x < X \leq x+h)$ tiende al conjunto vacío. Concluimos entonces que, cuando $h \to 0$ con $h > 0$, $F(x+h) \to F(x) + P(\emptyset) = F(x)$. ■

Recíprocamente, toda función $F(x) : \mathbb{R} \to [0,1]$ que cumpla las cuatro propiedades anteriores (sin tener una variable aleatoria de por medio) se le llama función de distribución. La propiedad (c) significa que $F(x)$ es una función monótona no decreciente, mientras que la propiedad (d) establece que $F(x)$ es una función continua por la derecha. Observe que a partir de la demostración desarrollada para la propiedad (c), tenemos que

$$P(x_1 < X \leq x_2) = F(x_2) - F(x_1),$$

en donde X es una variable aleatoria con función de distribución $F(x)$ y $x_1 \leq x_2$.

EJEMPLO 1.55. Sea X una variable aleatoria con función de distribución

$$F(x) = \begin{cases} 0 & \text{si } x < -1, \\ 1/3 & \text{si } -1 \leq x < 0, \\ 2/3 & \text{si } 0 \leq x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$

Como un ejemplo del cálculo de probabilidades usando la función de distribución, verifique el lector los siguientes resultados:

a) $P(X \leq 1) = 1$.

b) $P(X > 0) = 1/3$.
c) $P(0 < X \leq 1) = 1/3$.
d) $P(X = 0) = 1/3$.

Para concluir esta sección mencionaremos que el término distribución o distribución de probabilidad de una v.a. se refiere a cualquiera de las dos funciones: a la función de probabilidad (o de densidad) $f(x)$ o a la función de distribución $F(x)$.

$$\begin{array}{c}\text{Distribución}\\ \text{de}\\ \text{probabilidad}\end{array} = \begin{cases}\text{Función de probabilidad}\\ \text{(o de densidad) } f(x)\\ \\ \text{Función de distribución } F(x)\end{cases}$$

7. Esperanza, varianza, momentos

Todos los seres humanos tenemos características numéricas que nos identifican y nos distinguen de otras personas, por ejemplo, la edad, la estatura, la talla, el peso, etc. Si pudiéramos considerar la totalidad de todos estos números para una persona en particular, la identificaríamos de manera única. Algo similar sucede con las variables aleatorias. En esta sección estudiaremos algunas características numéricas asociadas a las variables aleatorias que, bajo ciertas condiciones, las identifican de manera única.

Esperanza

La esperanza de una variable aleatoria X es un número denotado por $E(X)$ y que se calcula como sigue: si X es discreta con valores $x_0, x_1, \ldots$, entonces

$$\mathrm{E}(X) = \sum_{i=0}^{\infty} x_i\, P(X = x_i),$$

suponiendo que esta suma es absolutamente convergente, es decir, la suma de los valores absolutos es convergente. El número de sumandos puede ser finito o infinito, dependiendo del conjunto de valores de la variable aleatoria. Si X es continua con función de densidad $f(x)$, entonces la esperanza es

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x)\, dx,$$

suponiendo que esta integral es absolutamente convergente, esta hipótesis significa que la integral de los valores absolutos es convergente. Si la suma o integral anteriores no cumplen esta condición de convergencia absoluta, entonces se dice que la esperanza no existe. La esperanza de una variable aleatoria es entonces un número que indica el promedio ponderado de los diferentes valores que puede tomar la variable. A la esperanza se le conoce también con los nombre de: media, valor esperado o valor promedio. En general se usa la letra griega μ (mu) para denotarla. Hemos señalado que

la integral o suma arriba mencionados pueden no ser convergentes y en ese caso se dice que la variable aleatoria no tiene esperanza finita, en los ejercicios 147 y 148 pueden encontrarse ejemplos en donde se ilustra esta situación. La esperanza es uno de los conceptos más importantes en probabilidad y tiene un amplio uso en las aplicaciones y otras ramas de la ciencia. Para llevar a cabo el cálculo de esperanzas, a partir de este momento resultará muy útil conocer algunas fórmulas para llevar a cabo sumas y poseer un pleno manejo de las técnicas de integración. Mediante algunos ejemplos ilustraremos a continuación la forma de calcular esperanzas.

EJEMPLO 1.56 (Caso discreto). Sea X una variable aleatoria discreta con función de probabilidad dada por la siguiente tabla:

x	-1	0	1	2
$f(x)$	1/8	4/8	1/8	2/8

La esperanza de X es el número

$$\begin{aligned} E(X) &= \sum_x x\, f(x) \\ &= -1 \times 1/8 + 0 \times 4/8 + 1 \times 1/8 + 2 \times 2/8 \\ &= 1/2. \end{aligned}$$

Observe que la suma se efectúa para todos los valores de x indicados en la tabla, es decir: $-1, 0, 1$ y 2. También es instructivo observar que la esperanza no es necesariamente uno de los valores tomados por la variable aleatoria. En este ejemplo el valor $1/2$ nunca es tomado por la variable aleatoria pero es su valor esperado.

EJEMPLO 1.57 (Caso continuo). Considere la variable aleatoria continua X con función de densidad $f(x) = 2x$, para $x \in (0, 1)$, siendo cero fuera de este intervalo. La esperanza de X es

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_0^1 x\, 2x\, dx = \frac{2}{3}x^3 \Big|_0^1 = \frac{2}{3}.$$

Observe que la integral sólo es relevante en el intervalo $(0, 1)$, pues fuera de dicho intervalo la función de densidad se anula.

EJEMPLO 1.58. Considere una variable aleatoria continua X con función de densidad $f(x) = |x|$, para $-1 < x < 1$. Compruebe que la esperanza de esta variable aleatoria es cero.

EJEMPLO 1.59. Sea X una variable aleatoria discreta con posibles valores en el conjunto $\{1, 2, \ldots, n\}$ y tal que la probabilidad de que tome cualquiera de estos valores es $1/n$. Compruebe que la esperanza de X es $(n + 1)/2$.

Esperanza de una función de una variable aleatoria

En algunos casos es necesario calcular la esperanza de una función de una variable aleatoria, por ejemplo, si X es una variable aleatoria, entonces es claro que $Y = X^2$ es una función de X y es también una variable aleatoria. Si quisiéramos calcular la esperanza de Y según la definición tendríamos que calcular

$$E(Y) = \int_{-\infty}^{\infty} y\, f_Y(y)\, dy,$$

para lo cual se necesita encontrar primero la función de densidad de Y y ello en general no es fácil. El siguiente resultado es muy útil y nos dice la forma de calcular esta esperanza conociendo únicamente la función de densidad de X. A este resultado a veces se le refiere como el teorema del estadístico inconsciente.

PROPOSICIÓN 1.60. *Sea X una variable aleatoria continua y sea $g : \mathbb{R} \to \mathbb{R}$ una función tal que $g(X)$ es una variable con esperanza finita. Entonces*

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx. \tag{4}$$

En general, la demostración de este resultado es complicada, así es que la omitiremos y nos concentraremos en su uso y aplicación. El resultado está enunciado en el caso continuo pero también es válido en el caso discreto, en donde en lugar de la integral aparece una suma.

EJEMPLO 1.61. Calcularemos $E(Y)$ en donde $Y = X^2$ y X es la variable aleatoria continua del ejemplo anterior, es decir, con función de densidad $f(x) = 2x$ para $x \in (0, 1)$. Por la proposición anterior tenemos que

$$E(Y) = E(X^2) = \int_{-\infty}^{\infty} x^2\, f(x)\, dx = \int_0^1 2x^3\, dx = 1/2.$$

Ahora, como un ejercicio para el lector, intente encontrar la función de densidad de Y y calcule $E(Y)$ usando la definición de esperanza. El resultado debe ser nuevamente $1/2$.

EJEMPLO 1.62. Sea X una variable aleatoria con función de probabilidad dada por la tabla que aparece abajo.

x	-2	-1	0	1	2
$f_X(x)$	2/8	1/8	2/8	1/8	2/8

Nos interesa calcular la esperanza de la variable aleatoria $Y = X^2$. Un primer método consiste en calcular primero la función de probabilidad de la variable Y. Esta función es

y	0	1	4
$f_Y(y)$	2/8	2/8	4/8

Entonces usando la definición de esperanza para variables aleatorias discretas tenemos que

$$E(Y) = 0\,\frac{2}{8} + 1\,\frac{2}{8} + 4\,\frac{4}{8} = \frac{9}{4}.$$

Ahora calcularemos la misma esperanza pero usando (4). El resultado es el mismo.

$$E(Y) = (-2)^2\frac{2}{8} + (-1)^2\frac{1}{8} + (0)^2\frac{2}{8} + (1)^2\frac{1}{8} + (2)^2\frac{2}{8} = \frac{9}{4}.$$

He aquí algunas propiedades generales de la esperanza.

PROPOSICIÓN 1.63. *Sean X y Y con esperanza finita y sea c una constante. Entonces*

a) $E(c) = c$.
b) $E(c\,X) = c\,E(X)$.
c) *Si* $X \geq 0$, *entonces* $E(X) \geq 0$.
d) $E(X + Y) = E(X) + E(Y)$.

La primera propiedad es evidente de demostrar pues si X es la variable aleatoria constante c, entonces por definición, $E(X) = c\,P(X = c) = c\,1 = c$. El segundo inciso se sigue directamente de la definición de esperanza pues tanto en el caso de la suma como en el caso de la integral, la constante c puede siempre colocarse fuera. El tercer inciso también es evidente pues cuando se cumple la hipótesis, en la integral o suma correspondiente solo aparecerán términos que son no negativos. La última propiedad, en cambio, no es sencilla de demostrar y aún en el caso discreto requiere de detalles técnicos que preferimos omitir. Observe que la segunda y cuarta propiedad establecen que la esperanza es lineal, es decir, separa sumas y también separa multiplicaciones por constantes.

Varianza

Vamos ahora a definir otra característica numérica asociada a las variables aleatorias llamada varianza. Se denota por $\text{Var}(X)$ y para una variable aleatoria discreta con valores $x_0, x_1, \ldots$ se define como sigue

$$\text{Var}(X) = \sum_{i=0}^{\infty} (x_i - \mu)^2 f(x_i),$$

en donde μ es la esperanza de X y bajo la hipótesis de que dicha suma es convergente. Por supuesto, cuando la variable aleatoria sólo toma un número finito de valores, la suma es finita. Observe que se necesita conocer la esperanza de X para calcular su varianza. En el caso continuo, para una variable aleatoria continua X con función de densidad $f(x)$ se define

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2\, f(x)\, dx.$$

Es interesante observar que la varianza puede escribirse en ambos casos (discreto y continuo) en una sola expresión como sigue

$$\text{Var}(X) = E\left[(X-\mu)^2\right].$$

Esto corresponde a la esperanza de una función de una variable aleatoria, siendo en este caso la función cuadrática $x \mapsto (x-\mu)^2$. La varianza es una medida del grado de dispersión de los diferentes valores tomados por la variable. Se le denota regularmente por la letra σ^2 (sigma cuadrada). A la raíz cuadrada positiva de la varianza, esto es σ, se le llama desviación estándar. Como en el caso de la esperanza, la anterior suma o integral puede no existir (no ser convergente), y en ese caso decimos que la variable aleatoria no tiene varianza finita. Veamos algunos ejemplos sencillos.

EJEMPLO 1.64. Calcularemos la varianza de la variable aleatoria discreta X con función de densidad dada por la siguiente tabla.

x	-1	0	1	2
$f(x)$	1/8	4/8	1/8	2/8

Recordemos primeramente que por cálculos previos, $\mu = 1/2$. Aplicando la definición de varianza tenemos que

$$\begin{aligned}
\text{Var}(X) &= \sum_x (x-\mu)^2\, f(x) \\
&= (-1-1/2)^2 \times 1/8 + (0-1/2)^2 \times 4/8 \\
&\quad +(1-1/2)^2 \times 1/8 + (2-1/2)^2 \times 2/8 \\
&= 1.
\end{aligned}$$

EJEMPLO 1.65. Calcularemos la varianza de la variable aleatoria continua X con función de densidad $f(x) = 2x$ para $x \in (0,1)$. En un cálculo previo habíamos encontrado que $\mu = 2/3$. Por lo tanto,

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x-\mu)^2\, f(x)\, dx = \int_0^1 (x-2/3)^2\, 2x\, dx = 1/18.$$

Ahora enunciamos algunas propiedades de la varianza.

PROPOSICIÓN 1.66. *Sean X y Y dos variables aleatorias, y sea c una constante. Entonces*

a) $\text{Var}(X) \geq 0$.
b) $\text{Var}(c) = 0$.
c) $\text{Var}(c\,X) = c^2\ \text{Var}(X)$.
d) $\text{Var}(X+c) = \text{Var}(X)$.
e) $\text{Var}(X) = E(X^2) - E^2(X)$.
f) *En general,* $\text{Var}(X+Y) \neq \text{Var}(X) + \text{Var}(Y)$.

DEMOSTRACIÓN. El inciso (a) es evidente a partir de la definición de varianza pues en ella aparece una suma o integral de términos no negativos. Para el inciso (b)

la constante c es una v.a. con un único valor, de modo que $E(c) = c$ y entonces $\text{Var}(X) = E(c-c)^2 = 0$. Para el inciso (c) tenemos que

$$\begin{aligned} \text{Var}(cX) &= E[cX - E(cX)]^2 \\ &= E[cX - cE(X)]^2 \\ &= c^2\, E[X - E(X)]^2 \\ &= c^2\, \text{Var}(X). \end{aligned}$$

El inciso (d) se sigue del siguiente análisis:

$$\text{Var}(X+c) = E[(X+c) - E(X+c)]^2 = E[X - E(X)]^2 = \text{Var}(X).$$

Para demostrar la propiedad (e) se desarrolla el cuadrado en la definición de varianza y se usa la propiedad de linealidad de la esperanza:

$$\begin{aligned} \text{Var}(X) &= E[X - E(X)]^2 \\ &= E[X^2 - 2XE(X) + E^2(X)] \\ &= E(X^2) - 2E(X)E(X) + E^2(X) \\ &= E(X^2) - E^2(X). \end{aligned}$$

Finalmente para demostrar la propiedad (f) es suficiente dar un ejemplo. Puede tomarse el caso $Y = X$, en general y por lo demostrado antes, no se cumple que $\text{Var}(2X) = 2\ \text{Var}(X)$. ■

De estas propiedades generales se obtiene en particular que la varianza es siempre una cantidad no negativa y que no cumple la propiedad de linealidad, pues en general no separa sumas y cuando aparecen constantes como factores, las constantes se separan de la varianza elevándolas al cuadrado.

EJEMPLO 1.67. Considere una variable aleatoria continua X con función de densidad dada por $f(x) = e^{-x}$, para $x > 0$. Usando la definición de esperanza y varianza puede comprobarse que $E(X) = 1$, y $\text{Var}(X) = 1$.

EJEMPLO 1.68. Sea X una variable aleatoria con varianza finita y sean a y b dos constantes. Usando las propiedades demostradas para la varianza puede verificarse la fórmula

$$\text{Var}(aX + b) = a^2\ \text{Var}(X).$$

Momentos

Finalmente definimos el n-ésimo momento de una variable aleatoria X, cuando existe, como el número $E(X^n)$ para cada $n = 1, 2, \ldots$ suponiendo que tal esperanza exista.

Así, los momentos de una variable aleatoria X son la colección de números:

$$\begin{array}{rl} E(X) & \text{Primer momento} \\ E(X^2) & \text{Segundo momento} \\ E(X^3) & \text{Tercer momento} \\ \cdots\cdots & \cdots\cdots \end{array}$$

Para variables aleatorias discretas el n-ésimo momento se calcula como sigue:

$$E(X^n) = \sum_{i=0}^{\infty} x_i^n \, f(x_i),$$

mientras que para variables aleatorias continuas la fórmula es la siguiente:

$$E(X^n) = \int_{-\infty}^{\infty} x^n \, f(x)\, dx.$$

Observe que el primer momento es simplemente la esperanza de la variable aleatoria y recordando la fórmula $\text{Var}(X) = E(X^2) - E^2(X)$, puede decirse ahora que la varianza es el segundo momento menos el primer momento al cuadrado. Existen interpretaciones conocidas para los primeros momentos pero es difícil encontrar alguna interpretación para todos ellos. Los momentos indican, sin embargo, alguna característica numérica de la variable aleatoria: el primer momento es el valor promedio, el segundo momento está relacionado con la dispersión de los valores de la variable aleatoria, el tercer momento está relacionado con la simetría de la correspondiente función de densidad. Esta situación de interpretación de los momentos de una variable aleatoria es muy similar a la situación de interpretación de las derivadas de una función: conocemos un interpretación mecánica o geométrica para las primeras derivadas de una función pero no conocemos un significado para todas y cada una de las derivadas. ¿Qué interpretación tiene la sexta derivada de una función? Bajo ciertas condiciones que establece el teorema de Taylor del cálculo diferencial e integral, a partir de las derivadas de una función en un punto se puede reconstruir la función misma. De manera análoga, bajo ciertas condiciones, a partir de todos los momentos de una variable aleatoria, se puede identificar plenamente a la variable aleatoria a través de su distribución.

EJEMPLO 1.69. Considere una variable aleatoria continua con función de densidad la que aparece abajo. No es difícil comprobar que el primer momento es cero y el segundo momento es $1/6$.

$$f(x) = \begin{cases} x+1 & \text{si } -1 < x < 0, \\ 1-x & \text{si } 0 \leq x < 1, \\ 0 & \text{otro caso.} \end{cases}$$

EJEMPLO 1.70. El n-ésimo momento de una variable aleatoria continua X con función de densidad $f(x) = x/2$, para $x \in (0, 2)$, es

$$E(X^n) = \frac{2^{n+1}}{n+2}.$$

Definimos también el n-ésimo momento central de X, cuando existe, como el número

$$E[(X-\mu)^n] \quad \text{para } n = 1, 2, \ldots$$

en donde $\mu = E(X)$. Observe que el segundo momento central es la varianza. Tenemos entonces que si X es una variable aleatoria discreta con valores $x_0, x_1, \ldots$, entonces el n-ésimo momento de X, si existe, se calcula como sigue:

$$E[(X-\mu)^n] = \sum_{i=0}^{\infty} (x_i - \mu)^n \, f(x_i) \, .$$

Para el caso continuo,

$$E[(X-\mu)^n] = \int_{-\infty}^{\infty} (x-\mu)^n \, f(x) \, dx.$$

EJEMPLO 1.71. Considere una variable aleatoria X con función de densidad

$$f(x) = \begin{cases} x & \text{si } 0 < x < 1, \\ 2-x & \text{si } 1 \leq x < 2, \\ 0 & \text{otro caso.} \end{cases}$$

Resolviendo la integral correspondiente puede encontrarse el n-ésimo momento central de esta variable aleatoria, dicho momento es

$$E(X-\mu)^n = \frac{1+(-1)^n}{(n+1)(n+2)}.$$

Cuantiles

Los cuantiles son otras características numéricas que pueden definirse para una distribución de probabilidad. Sea X una variable aleatoria con función de distribución $F(x)$ y sea p una probabilidad en el intervalo $(0, 1]$. Se dice que el número c_p es el cuantil p de X o de su distribución si es el número más pequeño que satisface

$$F(c_p) \geq p.$$

Esto significa que hasta el valor c_p la función de distribución acumula por lo menos una probabilidad de p. Véase la Figura 1.20. Cuando la distribución es continua, la desigualdad anterior puede escribirse en términos de una igualdad: el cuantil p es el número c_p más pequeño que cumple

$$F(c_p) = p.$$

Suelen utilizarse también las expresiones "c_p es el cuantil de orden p" o "c_p es el cuantil al $100p\,\%$", así tenemos por ejemplo, $c_{0.1}$ es el cuantil al $10\,\%$, $c_{0.2}$ es el cuantil al $20\,\%$, etc. Cuando la probabilidad p toma los valores 0.25, 0.50, 0.75 y 1, en lugar de cuantil se puede usar el término cuartil, de este modo la distribución de probabilidad se divide en cuatro cuartos de idéntica probabilidad cada uno. Véase la Figura 1.21. En este caso c_1 es finito pero no es difícil imaginar situaciones en donde es infinito.

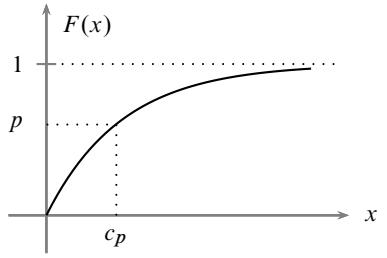


FIGURA 1.20

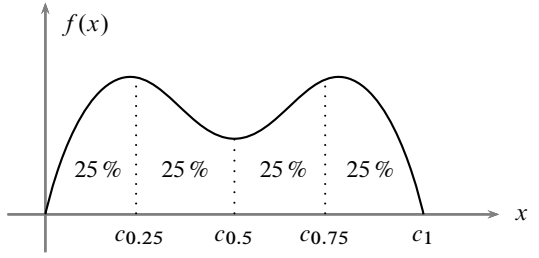


FIGURA 1.21

Se pueden ordenar los cuartiles de la siguiente forma:

$$\begin{array}{ccccccc} c_{0.25} & \leq & c_{0.50} & \leq & c_{0.75} & \leq & c_1 \\ \text{primer} & & \text{segundo} & & \text{tercer} & & \text{cuarto} \\ \text{cuartil} & & \text{cuartil} & & \text{cuartil} & & \text{cuartil} \end{array}$$

En particular, al cuantil al 50 % o segundo cuartil se le llama mediana de la variable aleatoria o de su distribución.

8. Distribuciones de probabilidad

Estudiaremos a continuación algunas distribuciones de probabilidad de variables aleatorias importantes. Estas distribuciones son modelos particulares para asignar probabilidades a subconjuntos de números reales. Empezaremos con las distribuciones de tipo discreto y continuaremos después con las de tipo continuo. Es importante señalar que presentaremos solamente una lista parcial de algunas distribuciones, aquellas posiblemente de mayor uso. En la última sección de este capítulo se explica la forma de simular en una computadora valores para algunas de estas distribuciones.

Distribución uniforme discreta

Decimos que una variable aleatoria X tiene una distribución uniforme discreta sobre el conjunto finito de números $\{x_1, \ldots, x_n\}$ si la probabilidad de que X tome cualquiera de estos valores es constante $1/n$. Esta distribución surge en espacios de probabilidad equiprobables, esto es, en situaciones en donde tenemos n resultados diferentes y todos ellos tienen la misma probabilidad de ocurrir. Los juegos de lotería son un ejemplo donde puede aplicarse esta distribución de probabilidad. Se escribe $X \sim \text{unif}\{x_1, x_2, \ldots, x_n\}$, en donde el símbolo "$\sim$" se lee "se distribuye como". La función de probabilidad de esta variable aleatoria es

$$f(x) = P(X = x) = \begin{cases} 1/n & \text{si } x = x_1, x_2, \ldots, x_n. \\ 0 & \text{otro caso.} \end{cases}$$

Es inmediato comprobar que la esperanza y la varianza para esta distribución se calculan del siguiente modo:

$$E(X) = \frac{1}{n}\sum_{i=1}^{n} x_i,$$

$$\text{y} \quad \text{Var}(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2.$$

EJEMPLO 1.72. La gráfica de la función de probabilidad de la distribución uniforme en el conjunto $\{1, 2, 3, 4, 5\}$ aparece en la Figura 1.22, junto con la correspondiente función de distribución. Cada salto en la función de distribución es de tamaño $1/5$.

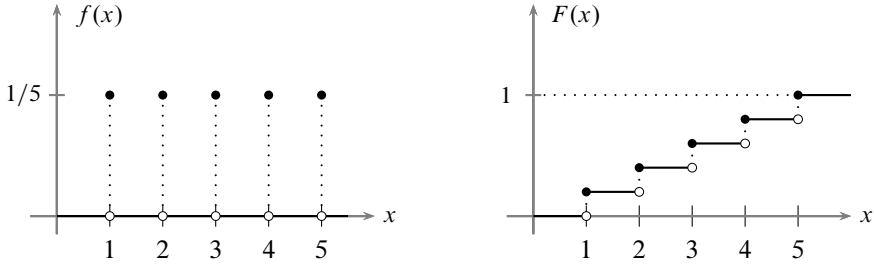


FIGURA 1.22

EJEMPLO 1.73. Al generar un número aleatorio en una computadora dentro del intervalo unitario $[0, 1]$ y debido a que la precisión de la computadora es necesariamente finita, se obtienen siempre valores dentro de un conjunto finito de elementos. Por ejemplo, si la precisión de la computadora es de dos decimales, entonces sólo se pueden generar los números : 0.00, 0.01, 0.02,. . ., 0.99, 1.00. La precisión de una computadora actual es claramente mucho mayor a la considerada pero siempre es finita

y la misma situación de imprecisión prevalece, es decir, en términos prácticos se tiene una distribución uniforme discreta al generar un valor al azar dentro del intervalo $[0, 1]$.

EJEMPLO 1.74. En Python se puede crear una lista de elementos y mediante la función predefinida *choice()* escoger un elemento al azar en la lista con distribución uniforme. El código es el siguiente:

```
>>> import random
>>> conjunto=[1,2,3,4,5]
>>> random.choice(conjunto)
3
```

Distribución Bernoulli

Un ensayo Bernoulli se define como aquel experimento aleatorio con únicamente dos posibles resultados llamados genéricamente: éxito y fracaso, con probabilidades respectivas p y $1-p$. Si se define la variable aleatoria X como aquella función que lleva el resultado éxito al número 1 y el resultado fracaso al número 0, entonces decimos que X tiene una distribución Bernoulli con parámetro $p \in (0, 1)$ y escribimos $X \sim \text{Ber}(p)$. La función de probabilidad se puede escribir de la siguiente forma:

$$P(X = x) = \begin{cases} p^x(1-p)^{1-x} & \text{si } x = 0, 1, \\ 0 & \text{otro caso.} \end{cases}$$

La gráfica de la función de probabilidad de esta distribución para $p = 0.7$ aparece en la Figura 1.23, junto con la correspondiente función de distribución. En este caso es muy sencillo verificar que $E(X) = p$ y $\text{Var}(X) = p(1-p)$. En la realización de todo experimento aleatorio siempre es posible preguntarnos por la ocurrencia o no ocurrencia de un evento cualquiera. Por ejemplo, ganar o no ganar en un juego de lotería, que llueva o no llueva hoy por la tarde, etc. Este es el esquema general donde surge esta distribución, que aunque sencilla, es importante y de amplia aplicación.

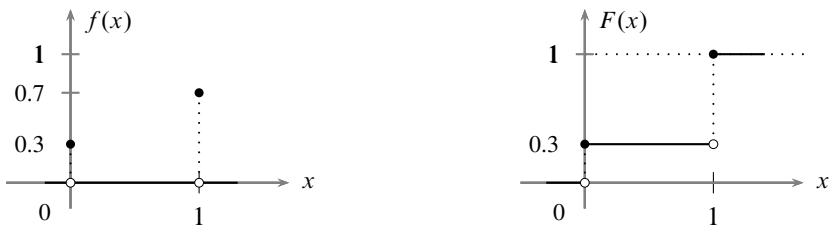


FIGURA 1.23

EJEMPLO 1.75. Considere el experimento aleatorio de lanzar una moneda al aire. Suponga que ω_1 y ω_2 son los dos resultados posibles, con probabilidades p y $1-p$, respectivamente. Sea X la variable aleatoria dada por $X(\omega_1) = 1$, y $X(\omega_2) = 0$.

Entonces X tiene distribución Ber(p). ¿Puede usted encontrar la distribución de la variable $1 - X$?

EJEMPLO 1.76. Sea Ω el espacio muestral de un experimento aleatorio cualquiera y sea $A \subseteq \Omega$ un evento tal que $P(A) = p > 0$. Sea X la variable aleatoria dada por la función indicadora del conjunto A, es decir, $X(\omega)$ es la función $1_A(\omega)$, la cual aparece definida abajo. Entonces X tiene distribución Ber(p).

$$X(\omega) = 1_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A, \\ 0 & \text{si } \omega \notin A. \end{cases}$$

Distribución binomial

Supongamos ahora que tenemos una serie de n ensayos independientes Bernoulli en donde la probabilidad de éxito en cualesquiera de estos ensayos es p. Si denotamos por E el resultado éxito y por F el resultado fracaso, entonces el espacio muestral consiste de todas las posibles sucesiones de longitud n de caracteres E y F. Usando el principio multiplicativo, es fácil ver que el conjunto Ω tiene 2^n elementos. Si ahora definimos la variable aleatoria X como aquella que cuenta el número de éxitos en cada una de estas sucesiones, esto es,

$$\begin{aligned} X(EE\cdots EE) &= n, \\ X(FE\cdots EE) &= n-1, \\ &\vdots \\ X(FF\cdots FF) &= 0, \end{aligned}$$

entonces tenemos que X puede tomar los valores $0, 1, 2, \ldots, n$ con las probabilidades que aparecen abajo. Decimos entonces que X tiene una distribución binomial con parámetros n y p y escribimos $X \sim \text{bin}(n, p)$. La función de probabilidad de esta variable aleatoria es

$$f(x) = P(X = x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{si } x = 0, 1, 2, \ldots, n. \\ 0 & \text{otro caso.} \end{cases}$$

Esta fórmula puede justificarse de la siguiente forma: deseamos obtener el evento de que en n ensayos Bernoulli aparezcan x éxitos y $n - x$ fracasos. Preliminarmente la probabilidad de este evento es el número

$$\underbrace{p\cdots p}_{x}\,\underbrace{(1-p)\cdots(1-p)}_{n-x} = p^x (1-p)^{n-x},$$

pero hemos colocado los x éxitos en los primeros x ensayos, tenemos entonces que multiplicar por las diferentes formas en que estos x éxitos pueden distribuirse en los n ensayos, este factor es el coeficiente binomial $\binom{n}{x}$. Después de algunos cálculos puede demostrarse que para esta distribución $E(X) = np$ y $\text{Var}(X) = np(1-p)$.

Los cálculos de las probabilidades de la distribución binomial, entre otras distribuciones que veremos más adelante, pueden involucrar muchas operaciones aritméticas. Para facilitar esta tarea se han creado varias calculadoras disponibles en internet que proveen los valores de las probabilidades solicitadas. Por ejemplo, en la página web *http://stattrek.com/tables/binomial.aspx* el lector puede encontrar una calculadora para encontrar las distintas probabilidades de la distribución binomial especificando el número de ensayos n, la probabilidad de éxito p y el número de éxitos x.

EJEMPLO 1.77. Cuando $n = 10$ ensayos, con probabilidad $p = 0.3$, se puede calcular, por ejemplo, $P(X = 2) = \binom{10}{2} (0.3)^2 (0.7)^{10-2} = 0.2334$. La gráfica de esta función de probabilidad con estos parámetros aparece en la Figura 1.24.

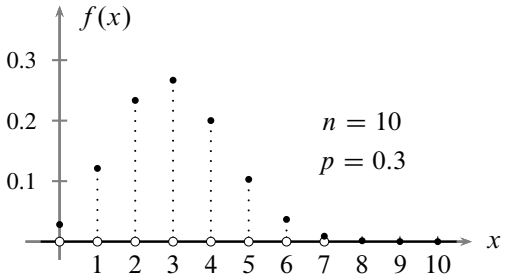


FIGURA 1.24

EJEMPLO 1.78. Un examen tiene diez preguntas y cada una de ellas tiene tres opciones como respuesta, siendo solamente una de ellas la correcta. Si un estudiante contesta cada pregunta al azar, ¿cuál es la probabilidad de que apruebe el examen?
Solución: si X denota el número de preguntas contestadas correctamente, entonces X tiene distribución bin(n, p) con $n = 10$ y $p = 1/3$. Suponiendo que la calificación mínima aprobatoria es 6, entonces la respuesta es

$$P(X \geq 6) = \sum_{x=6}^{10} \binom{10}{x} (1/3)^x (2/3)^{10-x} = 0.07656\,.$$

Esta probabilidad es sorprendentemente pequeña y por lo tanto la estrategia seguida por el estudiante para contestar el examen no parece ser muy buena.

Distribución geométrica

Supongamos que tenemos ahora una sucesión infinita de ensayos independientes Bernoulli, en cada uno de los cuales la probabilidad de éxito es p. Para cada una de estas sucesiones definimos la variable aleatoria X como el número de fracasos antes de

obtener el primer éxito. Por ejemplo,

$$\begin{aligned} X(FEFEFF\cdots) &= 1, \\ X(EFFEEE\cdots) &= 0, \\ X(FFFEFE\cdots) &= 3. \end{aligned}$$

Observamos que X puede tomar los valores $0, 1, 2, \ldots$ La probabilidad de que X tome el valor entero $x \geq 0$ es $p(1-p)^x$. Decimos entonces que X tiene una distribución geométrica con parámetro p y escribimos $X \sim \text{geo}(p)$ cuando su función de probabilidad es

$$f(x) = P(X = x) = \begin{cases} p\,(1-p)^x & \text{si } x = 0, 1, 2, \ldots \\ 0 & \text{otro caso.} \end{cases}$$

El nombre de esta distribución proviene del hecho de que cuando escribimos la suma de todas las probabilidades obtenemos una suma geométrica. Para esta distribución es posible demostrar que

$$\begin{aligned} E(X) &= \frac{1-p}{p}, \\ \text{Var}(X) &= \frac{1-p}{p^2}. \end{aligned}$$

La gráfica de la distribución geométrica de parámetro $p = 0.4$ se muestra en la Figura 1.25.

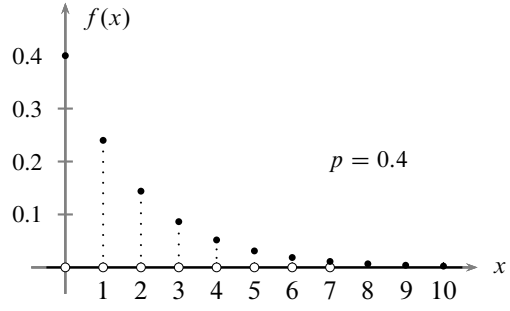


FIGURA 1.25

EJEMPLO 1.79. La inspección sucesiva de artículos hasta encontrar uno defectuoso, posiblemente en un proceso de control de calidad, puede modelarse usando una distribución geométrica.

EJEMPLO 1.80. Una persona participa cada semana con un boleto en un juego de lotería en donde la probabilidad de ganar el primer premio es $p = 10^{-6} = 1/1,000,000$. ¿Cuántos años en promedio debe esta persona participar en el juego

antes de obtener el primer premio?
Solución: si X denota el número de participaciones en el juego antes de obtener el primer premio, entonces X tiene distribución geométrica de parámetro $p = 10^{-6}$. Y por lo tanto la esperanza de la variable $X + 1$ es el número promedio de semanas que deben transcurrir para obtener el primer premio. Este número es

$$\frac{1-p}{p} + 1 = \frac{1}{p} = 10^6 = 1,000,000.$$

Lo que es equivalente a 19,230 años aproximadamente.

EJEMPLO 1.81. Sea X una variable aleatoria con distribución geo(p). Efectuando las sumas parciales de la función de probabilidad puede encontrarse la correspondiente función de distribución:

$$F(x) = P(X \leq x) = \sum_{u \leq x} f(u) = \begin{cases} 0 & \text{si } x < 0, \\ 1-(1-p)^1 & \text{si } 0 \leq x < 1, \\ 1-(1-p)^2 & \text{si } 1 \leq x < 2, \\ \dots & \dots \\ 1-(1-p)^{k+1} & \text{si } k \leq x < k+1, \\ \dots & \dots \end{cases}$$

En algunos textos se define la distribución geométrica contando el número de ensayos (no el de fracasos) antes del primer éxito. En este caso la variable es $1 + X$, es decir, la distribución se desplaza hacia la derecha una unidad, la esperanza es ahora $1/p$ y la varianza permanece constante $(1-p)/p^2$.

Distribución Poisson

Supongamos que deseamos observar el número de ocurrencias de un cierto evento dentro de un intervalo de tiempo dado, por ejemplo, el número de clientes que llegan a un cajero automático durante la noche, o tal vez deseamos registrar el número de accidentes que ocurren en cierta avenida durante todo un día. Para modelar este tipo de situaciones podemos definir la variable aleatoria X como el número de ocurrencia de este evento en el intervalo de tiempo dado. Es claro entonces que X puede tomar los valores $0, 1, 2, \dots$, y en principio no ponemos una cota superior para el número de observaciones del evento. Adicionalmente supongamos que conocemos la tasa media de ocurrencia del evento de interés, que denotamos por la letra λ (lambda). El parámetro λ es positivo y se interpreta como el número promedio de ocurrencias del evento por unidad de tiempo. La probabilidad de que la variable aleatoria X tome un valor entero $x \geq 0$ se definirá a continuación. Decimos que X tiene una distribución Poisson con parámetro $\lambda > 0$, y escribimos $X \sim \text{Poisson}(\lambda)$ cuando

$$P(X = x) = \begin{cases} e^{-\lambda}\,\dfrac{\lambda^x}{x!} & \text{si } x = 0, 1, 2, \dots \\ 0 & \text{otro caso.} \end{cases}$$

Puede demostrarse que la función $f(x)$ arriba definida es efectivamente una función de probabilidad para cada valor de $\lambda > 0$ fijo. Para ello se debe recordar el desarrollo de la serie de Taylor de la función e^x alrededor de $x = 0$. La forma de esta función se muestra en la Figura 1.26 cuando $\lambda = 2$. Después de algunos cálculos sencillos puede también comprobarse que $E(X) = \lambda$ y $\text{Var}(X) = \lambda$.

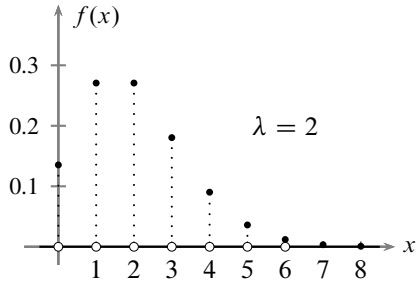


FIGURA 1.26

EJEMPLO 1.82. En promedio se reciben 2 peticiones de acceso a una página web durante un minuto cualquiera. Utilice el modelo Poisson para calcular la probabilidad de que en un minuto dado:

a) nadie solicite acceso a la página.
b) se reciban mas de dos peticiones.

Solución: sea X el número de peticiones por minuto y supongamos que X tiene distribución Poisson(λ) con $\lambda = 2$. Para el primer inciso se tiene que

$$P(X = 0) = e^{-2}\,\frac{2^0}{0!} = 0.135\,.$$

Para el segundo inciso,

$$\begin{aligned} P(X > 2) = 1 - P(X \leq 2) &= 1 - (P(X = 0) + P(X = 1) + P(X = 2)) \\ &= 1 - e^{-2}\,(\frac{2^0}{0!} + \frac{2^1}{1!} + \frac{2^2}{2!}\,) \\ &= 0.323\,. \end{aligned}$$

Un aspecto interesante de la distribución Poisson es que puede obtenerse como límite de la distribución binomial del siguiente modo.

PROPOSICIÓN 1.83. *Sea X con distribución* bin$(n\,,\,p)$ *y suponga que el número de ensayos n crece a infinito y la probabilidad de éxito p decrece a cero de tal forma que el producto np se mantiene constante igual a λ. Entonces para cualquier entero $x \geq 0$,*

$$\lim_{n\to\infty} P(X = x) = e^{-\lambda}\,\frac{\lambda^x}{x!}.$$

Este resultado puede comprobarse usando la fórmula de Stirling (ver Apéndice) y establece que cuando X tiene distribución bin(n, p) y hacemos tender n a infinito y p a cero de tal forma que el producto np se mantiene constante igual a λ, entonces la variable aleatoria X adquiere la distribución Poisson con parámetro λ. Este resultado sugiere que cuando n es grande, la distribución binomial puede ser aproximada mediante la distribución Poisson de parámetro $\lambda = np$. Esto es particularmente útil pues el cálculo de probabilidades de la distribución binomial involucra el cálculo de factoriales y ello puede ser computacionalmente demandante. El siguiente ejemplo ilustrará esta situación.

EJEMPLO 1.84. En promedio uno de cada 100 focos producido por una máquina es defectuoso. Use la distribución Poisson para estimar la probabilidad de encontrar 5 focos defectuosos en un lote de 1000 focos.
Solución: sea X el número de focos defectuosos en el lote de 1000 focos. Entonces X tiene distribución bin(n, p) con $n = 1000$ y $p = 1/100$. Nos piden calcular

$$P(X = 5) = \binom{1000}{5}(1/100)^5(99/100)^{995} = 0.0374531116\,.$$

Usando la aproximación Poisson, con $\lambda = np = 1000/100 = 10$,

$$P(X = 5) \approx e^{-10}\frac{10^5}{5!} = 0.0379841747\,.$$

Hemos definido a la variable aleatoria Poisson como el número de ocurrencias de un cierto evento dentro de un intervalo de tiempo dado, supongamos de longitud unitaria, $[0, 1]$. Suponga ahora que nos interesa observar las ocurrencias del evento en un intervalo de longitud diferente, por ejemplo $[0, t]$, con $t > 0$. Tal conteo de ocurrencias también sigue una distribución Poisson pero esta vez de parámetro λt. Por ejemplo, si $t = 2$, entonces el número de ocurrencias del evento en el intervalo $[0, 2]$ tiene distribución Poisson(2λ). El siguiente ejemplo ilustra esta situación.

EJEMPLO 1.85. El número de aviones que llegan a un aeropuerto internacional tiene una distribución Poisson con una frecuencia de 3 aviones cada 10 minutos. Es decir, la unidad de medición del tiempo son diez minutos. Entonces

a) La probabilidad de que no llegue ningún avión en un periodo de 20 minutos (dos unidades de tiempo) es $P(X = 0)$ con $\lambda = 3\,(2) = 6$.
b) La probabilidad de que llegue sólo un avión en el minuto siguiente es $P(X = 1)$ con $\lambda = 3\,(1/10) = 3/10$.
c) La probabilidad de que lleguen dos o mas aviones en un periodo de 15 minutos es $P(X \geq 2)$ con $\lambda = 3\,(1.5) = 4.5$.

Distribución binomial negativa

Consideremos nuevamente el experimento de llevar a cabo una sucesión de ensayos independientes Bernoulli, en cada uno de los cuales la probabilidad de éxito es p. Sea

r un parámetro fijo que toma alguno de los valores $1, 2, \ldots$. Si la variable aleatoria X cuenta el número de fracasos antes de obtener el r-ésimo éxito, entonces decimos que X tiene una distribución binomial negativa con parámetros r y p, y escribimos $X \sim \text{binneg}(r, p)$. Es claro que la variable X puede tomar los valores $0, 1, 2, \ldots$ con probabilidades como indica su función de probabilidad dada por

$$f(x) = P(X = x) = \begin{cases} \binom{r+x-1}{x} p^r (1-p)^x & \text{si } x = 0, 1, 2, \ldots \\ 0 & \text{otro caso.} \end{cases}$$

En esta fórmula aparece el término p^r pues la sucesión de ensayos Bernoulli no concluye sino hasta obtener r éxitos. Podemos tener un número variable de fracasos, de ahí el término $(1-p)^x$, y finalmente el factor $\binom{r+x-1}{x}$ que nos dice las diferentes formas en que los r éxitos pueden aparecer en los $r + x - 1$ ensayos realizados antes del último, que necesariamente fue un éxito. La gráfica de esta función aparece en la Figura 1.27 cuando los valores de los parámetros son $r = 3$ y $p = 0.2$.

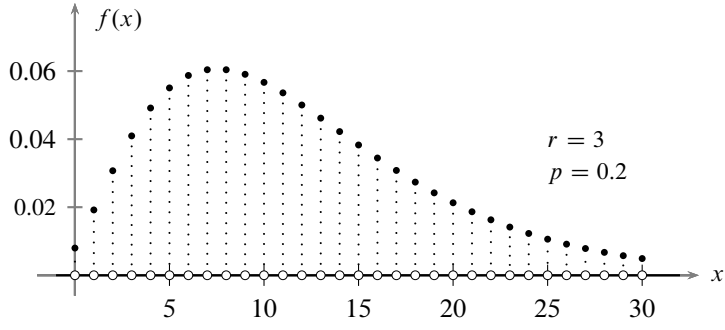


FIGURA 1.27

Es claro que esta distribución es una generalización de la distribución geométrica, la cual se obtiene tomando $r = 1$. Se puede además demostrar que $E(X) = r(1-p)/p$ y $\text{Var}(X) = r(1-p)/p^2$.

EJEMPLO 1.86. Se lanza repetidas veces una moneda honesta cuyos dos resultados son cara y cruz. ¿Cuál es la probabilidad de obtener la tercera cruz en el quinto lanzamiento? Solución: sea X el número de caras (fracasos) antes de obtener la tercera cruz. Entonces $X \sim \text{binneg}(r, p)$ con $r = 3$, $p = 1/2$, y nos preguntan $P(X = 2)$. Por lo tanto, $P(X = 2) = \binom{4}{2}(1/2)^5 = 6/32 = 0.1875$.

El nombre de la distribución binomial negativa surge del siguiente hecho: la definición de coeficiente binomial puede extenderse para cualquier número real a y

cualquier entero natural x de la siguiente manera, se define

$$\binom{a}{x} = \frac{a(a-1)\cdots(a-x+1)}{x!}.$$

Puede entonces demostrarse la siguiente identidad, de donde adquiere su nombre esta distribución.

$$\binom{n+x-1}{x} = (-1)^x \binom{-n}{x}.$$

Distribución hipergeométrica

Supongamos que tenemos un conjunto de N objetos de los cuales K son de una primera clase y $N-K$ son de una segunda clase. Véase la Figura 1.28. Supongamos que de este conjunto tomamos una muestra aleatoria de tamaño n, la muestra es sin reemplazo y el orden de los objetos seleccionados no importa.

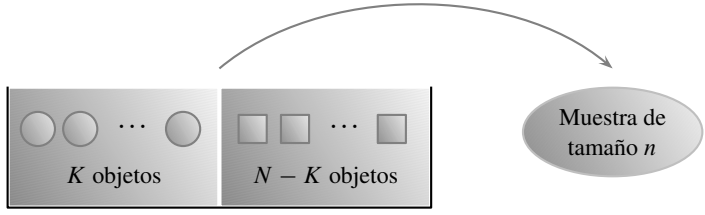


FIGURA 1.28

El espacio muestral de este experimento consiste de todas las posibles muestras de tamaño n que se pueden obtener del conjunto mayor de tamaño N. La cardinalidad del espacio muestral es $\binom{N}{n}$. Si definimos a la variable aleatoria X como el número de objetos de la primera clase contenidos en la muestra seleccionada, entonces X puede tomar los valores enteros x tales que

$$\text{máx}\{0, n-(N-K)\} \leq x \leq \text{mín}\{n, K\}.$$

en donde $\text{mín}\{n, K\}$ es la cantidad máxima de objetos de la primera clase que pueden tenerse en la muestra y $\text{máx}\{0, n-(N-K)\}$ es la cantidad mínima. Cuando $n \leq K$ y al mismo tiempo $n \leq N-K$, los posible valores para x son

$$0 \leq x \leq n.$$

Decimos entonces que X tiene una distribución hipergeométrica con parámetros N, K y n, y escribimos $X \sim \text{hipergeo}(N, K, n)$ si la probabilidad de que X tome alguno de los valores enteros arriba indicados está dada por

$$f(x) = P(X = x) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}.$$

El término $\binom{K}{x}$ nos dice las diferentes formas en que se pueden escoger x de los K objetos de la primera clase, y el término $\binom{N-K}{n-x}$ es nuevamente las diferentes formas de escoger $n-x$ objetos de los $N-K$ objetos de la segunda clase. De esta forma hemos usado el principio multiplicativo para obtener el número total de muestras diferentes en donde x objetos son de la primera clase y $n-x$ objetos son de la segunda clase. La gráfica de esta función de probabilidad para ciertos valores de los parámetros aparece en la Figura 1.29.

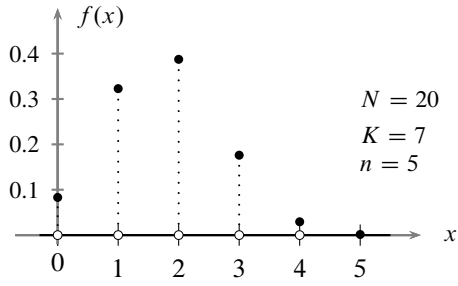


FIGURA 1.29

Para una variable aleatoria X con distribución hipergeo(N, K, n) tal que el tamaño de la muestra n es menor o igual a K y también menor o igual a $N-K$ (esto equivale a poder tener muestras en donde solo aparezca un solo tipo de objetos), puede demostrarse que

$$\begin{aligned} E(X) &= \frac{nK}{N}, \\ \text{Var}(X) &= n\frac{K}{N}\frac{N-K}{N}\frac{N-n}{N-1}. \end{aligned}$$

Con esto concluimos la revisión de algunas distribuciones de probabilidad de tipo discreto. Ahora estudiaremos algunas distribuciones de probabilidad de variables aleatorias continuas. No construiremos estas distribuciones a partir de experimentos aleatorios particulares como en el caso de algunas de las distribuciones discretas, mas bien, las definiremos sin mayor justificación. En algunos casos mostraremos la forma de obtener estas distribuciones a partir de considerar ciertas funciones de variables aleatorias conocidas. Algunos usos y aplicaciones de estos modelos serán mostrados en la segunda parte del texto correspondiente al tema de estadística.

Distribución uniforme continua

Decimos que una variable aleatoria X tiene una distribución uniforme continua en el intervalo (a, b), y escribimos $X \sim \text{unif}(a, b)$, cuando su función de densidad es

$$f(x) = \begin{cases} \dfrac{1}{b-a} & \text{si } x \in (a, b), \\ 0 & \text{otro caso.} \end{cases}$$

La gráfica general de esta función se muestra en la Figura 1.30(a), y es evidente que se trata de una función de densidad pues es no negativa e integra uno. En este caso es muy fácil encontrar la correspondiente función de distribución y ésta se muestra en la Figura 1.30(b). Los parámetros de esta distribución son los números a y b. Es fácil verificar que

$$\begin{aligned} E(X) &= (a+b)/2, \\ \text{y} \quad \text{Var}(X) &= (b-a)^2/12. \end{aligned}$$

Observe que la esperanza corresponde al punto medio del intervalo (a, b). Además la varianza o dispersión crece cuando a y b se alejan uno del otro, y por el contrario, cuando estos parámetros estan muy cercanos, la varianza es pequeña. Esta distribución es una de las más sencillas y sea usa naturalmente para cuando no se conoce mayor información de la variable aleatoria de interés, excepto que toma valores continuos dentro de cierto intervalo.

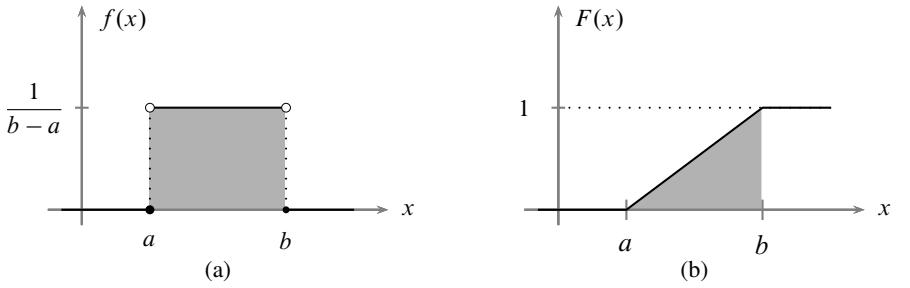


FIGURA 1.30. Distribución uniforme(a, b).

EJEMPLO 1.87. En el experimento aleatorio teórico de generar un número al azar X en el intervalo unitario $(0, 1)$ se considera regularmente que X tiene distribución uniforme en dicho intervalo.

EJEMPLO 1.88. Integrando la función de densidad uniforme en el intervalo (a, b) desde menos infinito hasta un punto x cualquiera, puede encontrarse la función de

distribución, la cual tiene la siguiente expresión y cuya gráfica se muestra en la Figura 1.30 (b).

$$F(x) = \begin{cases} 0 & \text{si } x < a, \\ \dfrac{x-a}{b-a} & \text{si } a \leq x < b, \\ 1 & \text{si } x \geq b. \end{cases}$$

Distribución exponencial

Decimos que una variable aleatoria continua X tiene una distribución exponencial con parámetro $\lambda > 0$, y escribimos $X \sim \exp(\lambda)$, cuando su función de densidad es

$$f(x) = \begin{cases} \lambda\, e^{-\lambda x} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

La gráfica de esta función cuando el parámetro λ toma el valor particular 3 se muestra en la Figura 1.31(a). La correspondiente función de distribución aparece a su derecha. Es muy sencillo verificar que la función $f(x)$ arriba definida es efectivamente una función de densidad para cualquier valor del parámetro $\lambda > 0$. Se trata pues de una variable aleatoria continua con conjunto de valores el intervalo $(0, \infty)$. Aplicando el método de integración por partes puede también comprobarse que $E(X) = 1/\lambda$, y $\text{Var}(X) = 1/\lambda^2$. Esta distribución se usa para modelar tiempos de espera para la ocurrencia de un cierto evento.

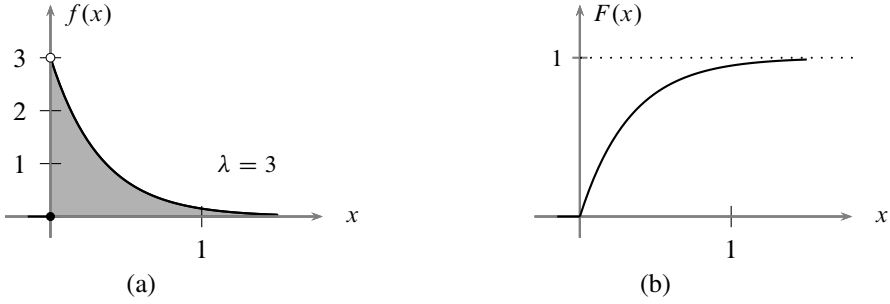


FIGURA 1.31. Distribución exponencial(λ).

EJEMPLO 1.89. Suponga que el tiempo en minutos que un usuario cualquiera permanece revisando su correo electrónico sigue una distribución exponencial de parámetro $\lambda = 1/5$. Calcule la probabilidad de que un usuario cualquiera permanezca conectado al servidor de correo

a) menos de un minuto.
b) mas de una hora.

Solución. Sea X el tiempo de conección al servidor de correo. Para el primer inciso tenemos que

$$P(X < 1) = \int_0^1 1/5\, e^{-x/5}\, dx = 0.181\,.$$

Para el segundo inciso,

$$P(X > 60) = \int_{60}^{\infty} 1/5\, e^{-x/5}\, dx = 0.0000061\,.$$

EJEMPLO 1.90. Sea X una variable aleatoria con distribución $\exp(\lambda)$. Integrando la función de densidad desde menos infinito hasta una valor arbitrario x se encuentra la función de distribución. La gráfica de esta función se muestra en la Figura 1.31(b) y tiene la siguiente expresión.

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

Distribución gama

La variable aleatoria continua X tiene una distribución gama con parámetros $n > 0$ y $\lambda > 0$, y escribimos $X \sim \text{gama}(n, \lambda)$, si su función de densidad es

$$f(x) = \begin{cases} \dfrac{(\lambda x)^{n-1}}{\Gamma(n)}\, \lambda e^{-\lambda x} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

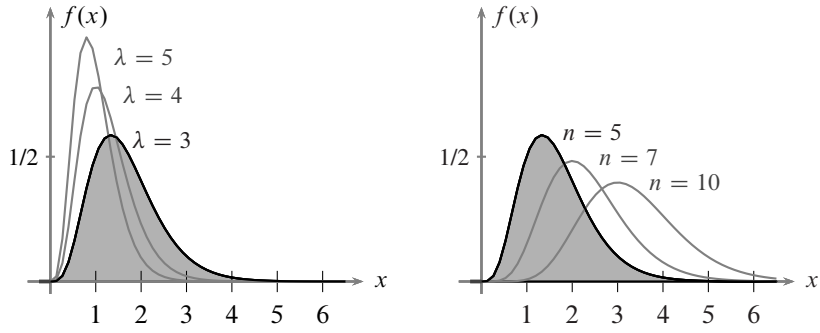
La gráfica de esta función de densidad para varios valores de los parámetros se muestra en la Figura 1.32. A partir de la expresión anterior puede observarse que los posibles valores para una variable aleatoria con esta distribución son aquellos números dentro del intervalo $(0, \infty)$.

En la expresión anterior aparece el término $\Gamma(n)$. Esta es la función gama que se define como la siguiente integral

$$\Gamma(n) = \int_0^{\infty} t^{n-1} e^{-t}\, dt,$$

para cualquier número real n tal que esta integral sea convergente. Esta función no es el tipo de funciones a las que estamos acostumbrados en los cursos de matemáticas elementales, en donde regularmente se conoce la expresión exacta de una cierta función y se utiliza esta expresión para evaluarla. En este caso, para evaluar la función gama es necesario substituir el valor de n en el integrando y efectuar la integral infinita. Afortunadamente no evaluaremos esta integral para cualquier valor de n, sólo para algunos pocos valores, principalmente enteros, y nos ayudaremos de las siguientes propiedades que no son difíciles de verificar:

a) $\Gamma(n+1) = n\,\Gamma(n)$.
b) $\Gamma(n+1) = n!$ si n es entero.



(a) $n = 5$ (b) $\lambda = 3$

FIGURA 1.32. Función de densidad gama(n, λ).

c) $\Gamma(2) = \Gamma(1) = 1$.
d) $\Gamma(1/2) = \sqrt{\pi}$.

El nombre de la distribución gama se deriva del hecho de que en su definición aparece la función gama. Observemos además que la distribución exponencial es un caso particular de la distribución gama: en efecto, si en la distribución gama tomamos el parámetro n igual a 1, obtenemos la distribución exponencial con parámetro λ. Resolviendo un par de integrales se puede demostrar que $E(X) = n/\lambda$, y $\text{Var}(X) = n/\lambda^2$. Cuando el parámetro n es un número natural, la distribución gama(n, λ) puede construirse de la forma como indica la proposición que aparece abajo. El concepto de independencia de variables aleatorias, el cual es una extensión del concepto de independencia de eventos, será será estudiado más adelante.

PROPOSICIÓN 1.91. *Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes cada una de ellas con distribución* exp(λ). *Entonces la variable aleatoria suma $X_1 + X_2 + \cdots + X_n$ tiene distribución* gama(n, λ).

Así, una variable aleatoria con distribución gama(n, λ), cuando n es un entero positivo, puede interpretarse como el tiempo acumulado de n tiempos de espera exponenciales independientes uno seguido del otro. En este caso a la distribución gama se le conoce también con el nombre de distribución Erlang(n, λ).

Distribución beta

Decimos que la variable aleatoria continua X tiene una distribución beta con parámetros $a > 0$ y $b > 0$, y escribimos $X \sim \text{beta}(a, b)$, cuando su función de densidad es

$$f(x) = \begin{cases} \dfrac{1}{B(a,b)}\, x^{a-1}\,(1-x)^{b-1} & \text{si } x \in (0,1), \\ 0 & \text{otro caso.} \end{cases}$$

El término $B(a,b)$ se conoce como la función beta, y de allí adquiere su nombre esta distribución. La función beta se define como sigue:

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx,$$

para números reales $a > 0$ y $b > 0$. Esta función está relacionada con la función gama, antes mencionada, a través de la identidad:

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

Véase la sección de ejercicios para una lista de propiedades de esta función. Para la distribución beta(a,b) se tiene que $E(X) = a/(a+b)$, y $\text{Var}(X) = ab/((a+b+1)(a+b)^2)$. La gráfica de la función de densidad para la distribución beta se muestra en la Figura 1.33, para distintos valores de los parámetros a y b.

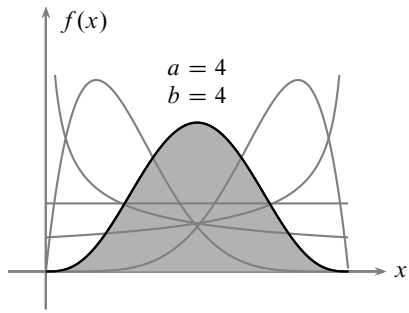


FIGURA 1.33. Función de densidad beta(a,b).

La distribución beta puede obtenerse a partir de la distribución gama como indica el siguiente resultado.

PROPOSICIÓN 1.92. *Sean X y Y dos variables aleatorias independientes con distribución* gama(a,λ) *y* gama(b,λ)*, respectivamente. Entonces la variable aleatoria $X/(X+Y)$ tiene distribución* beta(a,b).

Distribución Weibull

La variable aleatoria continua X tiene una distribución Weibull con parámetros $\alpha > 0$ y $\lambda > 0$ si su función de densidad está dada por la siguiente expresión

$$f(x) = \begin{cases} e^{-(\lambda x)^\alpha}\lambda\alpha\,(\lambda x)^{\alpha-1} & \text{si } x > 0, \\ 0 & \text{otro caso.} \end{cases}$$

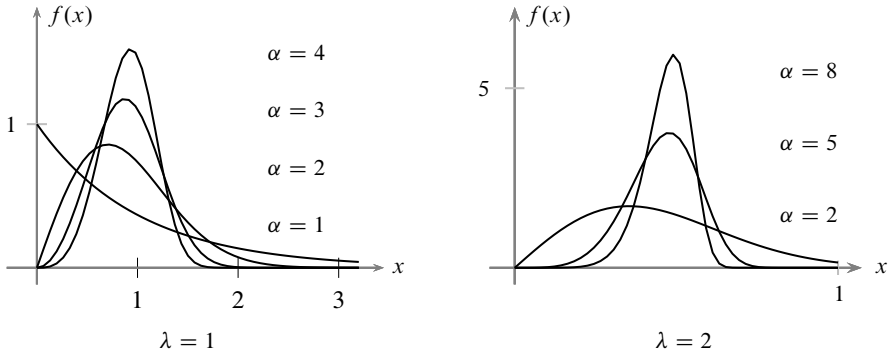


FIGURA 1.34. Función de densidad Weibull(α, λ).

A la constante α se le llama parámetro de forma y a λ se le llama parámetro de escala. Se escribe $X \sim \text{Weibull}(\alpha, \lambda)$. La gráfica de la función de densidad para varios valores de los parámetros se encuentra en la Figura 1.34.

La distribución Weibull se ha utilizado en estudios de confiabilidad y durabilidad de componentes electrónicos y mecánicos. El valor de una variable aleatoria X con esta distribución puede interpretarse como el tiempo de vida útil que tiene un componente en particular. Después de algunos cálculos puede encontrarse que la esperanza y varianza de este tiempo de vida útil son

$$\begin{aligned} E(X) &= \frac{1}{\lambda}\,\Gamma(1+1/\alpha), \\ \text{y} \quad \text{Var}(X) &= \frac{1}{\lambda^2}\,(\Gamma(1+2/\alpha) - \Gamma^2(1+1/\alpha)). \end{aligned}$$

Por otro lado, llevando a cabo un cambio de variable puede demostrarse que la correspondiente función de distribución es

$$F(x) = \begin{cases} 1 - e^{-(\lambda x)^{\alpha}} & \text{si } x > 0, \\ 0 & \text{otro caso.} \end{cases}$$

Cuando el parámetro α toma el valor uno, la distribución Weibull se reduce a la distribución exponencial de parámetro λ. Como veremos en la siguiente sección, conocer una expresión de la función de distribución nos ayudará a aplicar un mecanismo sencillo para simular valores al azar de esta variable aleatoria.

Distribución normal

Esta es posiblemente la distribución de probabilidad de mayor importancia. Decimos que la variable aleatoria continua X tiene una distribución normal si su función de

densidad está dada por la siguiente expresión

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-(x-\mu)^2/2\sigma^2}, \qquad x \in \mathbb{R},$$

en donde $\mu \in \mathbb{R}$ y $\sigma > 0$ son dos parámetros. Escribimos entonces $X \sim \text{N}(\mu, \sigma^2)$. La gráfica de esta función de densidad tiene forma de campana como se puede apreciar en la Figura 1.35, en donde se muestra además el significado geométrico de los dos parámetros.

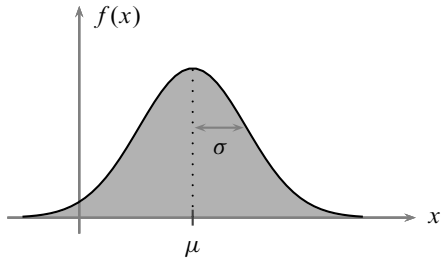


FIGURA 1.35. Función de densidad $\text{N}(\mu, \sigma^2)$.

Es claro que los posibles valores de una variable aleatoria con esta distribución es la totalidad de los números reales. Por otro lado, no es difícil demostrar que $E(X) = \mu$, y ello significa que la campana esta centrada en este valor, el cual puede ser negativo, positivo o cero. También puede demostrarse que $\text{Var}(X) = \sigma^2$, y que la distancia del punto μ a cualquiera de los dos puntos en donde la función tiene puntos de inflexión es σ, por lo tanto la campana se abre o se cierra de acuerdo a la magnitud de este parámetro.

Distribución normal estándar

Decimos que la variable aleatoria X tiene una distribución normal estándar si tiene una distribución normal con parámetros $\mu = 0$ y $\sigma^2 = 1$. En este caso la función de densidad se reduce a la expresión

$$f(x) = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}, \qquad x \in \mathbb{R}.$$

Es posible transformar una variable aleatoria normal no estándar en una estándar mediante la siguiente operación.

PROPOSICIÓN 1.93. *Sea X una variable aleatoria con distribución normal con parámetros μ y σ^2. Entonces la siguiente variable aleatoria tiene una distribución normal estándar.*

$$Z = \frac{X - \mu}{\sigma}. \tag{5}$$

Al procedimiento anterior se le conoce con el nombre de estandarización y bajo tal transformación se dice que la variable X ha sido estandarizada. Es común usar la letra Z para denotar una variable aleatoria con distribución normal estándar, y seguiremos nosotros también esa costumbre. Se deja como ejercicio al lector verificar que realmente la variable aleatoria Z tiene una distribución normal y que $E(Z) = 0$ y $\text{Var}(Z) = 1$. Este resultado parece muy modesto pero tiene una gran importancia operacional pues establece que el cálculo de las probabilidades de una variable aleatoria normal cualquiera se reduce al cálculo de las probabilidades para la normal estándar. Explicaremos ahora con más detalle esta situación: suponga que X es una variable aleatoria con distribución $\text{N}(\mu, \sigma^2)$ y que deseamos calcular, por ejemplo, la probabilidad de que X tome un valor en el intervalo (a, b), es decir, $P(a < X < b)$, para $a < b$ números dados. Tenemos entonces que

$$\begin{aligned} P(a < X < b) &= P(a - \mu < X - \mu < b - \mu) \\ &= P(\frac{a-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{b-\mu}{\sigma}) \\ &= P(\frac{a-\mu}{\sigma} < Z < \frac{b-\mu}{\sigma}). \end{aligned}$$

La igualdad de estas probabilidades es consecuencia de la igualdad de los eventos correspondientes. De esta forma una probabilidad que involucra a la variable X se ha reducido a una probabilidad que involucra a Z. De modo que únicamente necesitamos conocer las probabilidades de los eventos de Z para calcular las probabilidades de los eventos de la variable X que tiene parámetros arbitrarios.

Función de distribución $\text{N}(0, 1)$

Es común denotar a la función de distribución de una variable aleatoria con distribución normal estándar Z como $\Phi(x)$, es decir,

$$\boxed{\Phi(x) = P(Z \leq x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du,}$$

cuyo significado geométrico se muestra en la Figura 1.36(a). Seguramente el lector se sorprenderá al enterarse que, sin importar el método de integración que se utilice, no es posible resolver esta integral y encontrar una expresión exacta para $\Phi(x)$. Puede usted intentar resolver esta integral si lo desea y verificar la dificultad del problema. En la práctica lo que se hace es usar métodos numéricos para encontrar aproximaciones de $\Phi(x)$ para distintos valores de x. En la parte final del texto aparece una tabla con estos valores aproximados. Cada renglón de esta tabla corresponde a un valor de x hasta el primer dígito decimal, las distintas columnas corresponden al segundo dígito decimal. El valor que aparece en la tabla es $\Phi(x)$. Por ejemplo, el renglón marcado con 1.4 y la columna marcada con 0.05 corresponden al valor $x = 1.45$, tenemos entonces que $\Phi(1.45) = 0.9265$. Abajo aparecen algunos ejemplos que ilustran el uso de esta tabla. Observe además que para $x \geq 3.5$, la probabilidad $\Phi(x)$ es muy cercana a uno, es decir,

para esos valores de x la campana prácticamente ha decaido a cero en el lado derecho. Esto quiere decir que, con probabilidad casi idéndica a uno, los valores que toma una variable aleatoria normal estándar están comprendidos entre -3.5 y $+3.5$. Por otro lado, a partir del hecho de que si X tiene distribución normal estándar, entonces la variable $-X$ también tiene distribución normal estándar, puede demostrarse que

$$\Phi(-x) = 1 - \Phi(x).$$

Un argumento geométrico también puede utilizarse para darse cuenta de la validez de esta igualdad. En particular, este resultado ayuda a calcular valores de $\Phi(x)$ para x negativos en tablas de la distribución normal como la presentada al final del texto en donde sólo aparecen valores positivos para x.

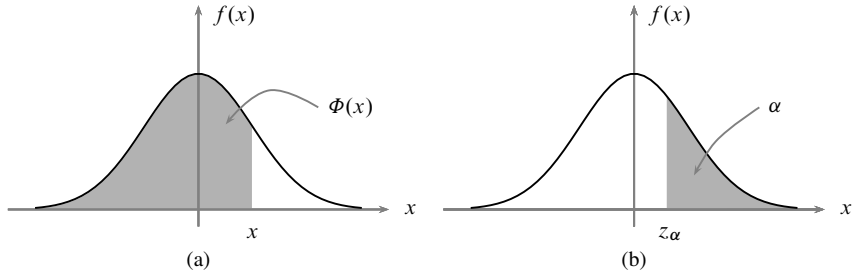


FIGURA 1.36. $\Phi(x)$ y notación z_α.

EJEMPLO 1.94. Use la tabla de la distribución normal estándar para comprobar que

a) $\Phi(1.65) = 0.9505$.
b) $\Phi(-1.65) = 0.0495$.
c) $\Phi(-1) = 0.1587$.

EJEMPLO 1.95. Use la tabla de la distribución normal estándar para encontrar el valor de x tal que

a) $\Phi(x) = 0.3$. Respuesta: $x = -0.53$.
b) $\Phi(x) = 0.75$. Respuesta: $x = 0.68$.

EJEMPLO 1.96. Sea X con distribución N(5, 10). Use el proceso de estandarización y la tabla de la distribución normal estándar para comprobar que

a) $P(X \leq 7) = 0.7357$.
b) $P(0 < X < 5) = 0.2357$.
c) $P(X > 10) = 0.0571$.

A continuación definiremos el número z_α, el cual será usado con regularidad en la segunda parte del texto.

NOTACIÓN 1.97. Para cada valor α en el intervalo $(0, 1)$, el número z_α denotará el punto en el eje real para el cual el área bajo la curva a la derecha de z_α es α. Esto es,

$$\Phi(z_\alpha) = 1 - \alpha.$$

El significado geométrico del número z_α se muestra en la Figura 1.36(b) y corresponde al cuantil $1 - \alpha$ de la distribución normal estándar.

EJEMPLO 1.98. Usando la tabla de la distribución normal estándar puede comprobarse que, de manera aproximada,

a) $z_{0.1} = 1.285$.
b) $z_{0.2} = 0.845$.

Distribución ji-cuadrada

Decimos que la variable aleatoria continua X tiene una distribución ji-cuadrada con n grados de libertad (n entero positivo), si su función de densidad está dada por la siguiente expresión:

$$f(x) = \begin{cases} \dfrac{1}{\Gamma(n/2)} \left(\dfrac{1}{2}\right)^{n/2} x^{n/2-1} e^{-x/2} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

Se trata de una variable aleatoria continua con posibles valores en el intervalo $(0, \infty)$. Esta distribución tiene un solo parámetro denotado aquí por la letra n, y al cual se le llama grados de libertad. A pesar de su aparente expresión complicada, no es difícil comprobar que $f(x)$ es efectivamente una función de densidad, y para ello se utiliza la definición de la función gama. La gráfica de esta función de densidad para varios valores del parámetro n aparece en la Figura 1.37.

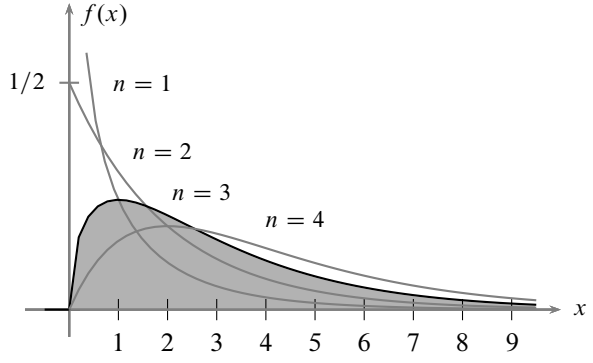


FIGURA 1.37. Función de densidad $\chi^2(n)$.

Escribiremos simplemente $X \sim \chi^2(n)$, en donde la letra griega χ se pronuncia "ji" o también "chi". Por lo tanto, la expresión "$\chi^2(n)$" se lee "ji cuadrada con n grados de libertad". Puede demostrarse sin mucha dificultad que

$$\begin{aligned} E(X) &= n, \\ \text{y} \quad \text{Var}(X) &= 2n. \end{aligned}$$

La distribución ji-cuadrada puede obtenerse como indican los varios resultados que a continuación enunciaremos y cuyas demostraciones omitiremos. En la segunda parte del texto se mostrará la utilidad y usos de estos resultados teóricos.

PROPOSICIÓN 1.99. *Si* $X \sim \text{N}(0,1)$, *entonces* $X^2 \sim \chi^2(1)$.

Es decir, el cuadrado de una variable aleatoria con distribución normal estándar tiene distribución ji-cuadrada con un grado de libertad. Por otro lado, el siguiente resultado establece que la suma de dos variables aleatorias independientes con distribución ji-cuadrada tiene distribución nuevamente ji-cuadrada con grados de libertad la suma de los grados de libertad de los sumandos.

PROPOSICIÓN 1.100. *Si* $X \sim \chi^2(n)$ *y* $Y \sim \chi^2(m)$ *son dos variables aleatorias independientes, entonces* $X + Y$ *tiene distribución* $\chi^2(n+m)$.

El resultado anterior puede extenderse al caso cuando se tienen varias variables aleatorias independientes con distribución χ^2. En particular, si $X_1, \ldots, X_n$ son variables independientes con distribución normal estándar, entonces la suma de los cuadrados $X_1^2 + \cdots + X_n^2$ tiene distribución $\chi^2(n)$. De este modo, si conocemos una forma de simular n valores al azar de la distribución normal estándar, la suma de los cuadrados de los números obtenidos será una observación de la distribución ji-cuadrada con n grados de libertad. Por último, mencionaremos el siguiente resultado que será utilizado en la sección de pruebas de hipótesis.

PROPOSICIÓN 1.101. *Sean* $X_1, \ldots, X_n$ *variables aleatorias independientes cada una de ellas con distribución* $\text{N}(\mu, \sigma^2)$. *Entonces*

$$\frac{(n-1)\,S^2}{\sigma^2} \sim \chi^2(n-1),$$

en donde $S^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2$ *y* $\bar{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$.

Distribución t

Decimos que la variable aleatoria continua X tiene una distribución t con n grados de libertad si su función de densidad está dada por

$$f(x) = \frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\;\Gamma(n/2)}\,(1 + x^2/n)^{-(n+1)/2}, \qquad x \in \mathbb{R}.$$

En tal caso se escribe $X \sim t(n)$. La gráfica de esta función de densidad aparece en la Figura 1.38. Es posible demostrar que $E(X) = 0$ para $n > 1$, y $\text{Var}(X) = n/(n-2)$ para $n > 2$. La distribución t se puede encontrar en los siguientes contextos:

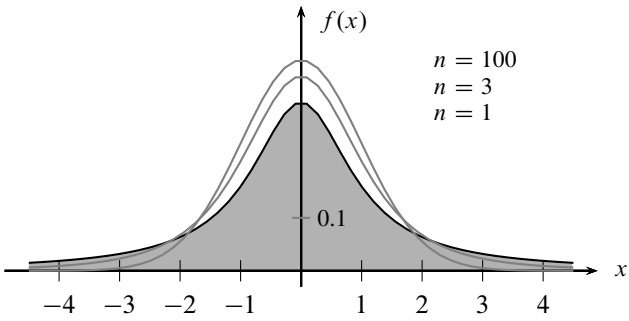


FIGURA 1.38. Función de densidad $t(n)$.

PROPOSICIÓN 1.102. *Si* $X \sim \text{N}(0,1)$ *y* $Y \sim \chi^2(n)$ *son independientes, entonces*

$$\frac{X}{\sqrt{Y/n}} \sim t(n).$$

El resultado anterior provee de un mecanismo para generar simulaciones de los valores que toma una variable aleatoria con distribución $t(n)$. Para ello se pueden generar n observaciones de la distribución normal estándar, y con ello conformar una observación de la distribución $\chi^2(n)$ como fue explicado antes. Se necesita una observación adicional de la distribución normal estándar que será el valor de X, según la fórmula de la proposición anterior, se hace el cociente indicado y el resultado será un valor de la distribución $t(n)$.

PROPOSICIÓN 1.103. *Sean* $X_1, \ldots, X_n$ *variables aleatorias independientes cada una de ellas con distribución* $\text{N}(\mu, \sigma^2)$. *Entonces*

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1),$$

en donde $\bar{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ *y* $S^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2$.

Distribución F

La variable aleatoria continua X tiene una distribución F de Fisher con parámetros $a > 0$ y $b > 0$ si su función de densidad está dada por la siguiente expresión:

$$f(x) = \begin{cases} \dfrac{\Gamma(\frac{a+b}{2})}{\Gamma(\frac{a}{2})\,\Gamma(\frac{b}{2})}\,(\frac{a}{b})^{a/2}\,x^{a/2-1}\,(1+\frac{a}{b}x)^{-(a+b)/2} & \text{si } x > 0, \\ 0 & \text{si } x \leq 0. \end{cases}$$

En este caso se escribe $X \sim \mathrm{F}(a, b)$. Una gráfica de esta función de densidad aparece en la Figura 1.39. Se puede demostrar que

$$E(X) = \frac{b}{b-2}, \qquad \text{si } b > 2,$$

$$\text{y} \quad \mathrm{Var}(X) = \frac{2b^2(a+b-2)}{a(b-2)^2(b-4)}, \qquad \text{si } b > 4.$$

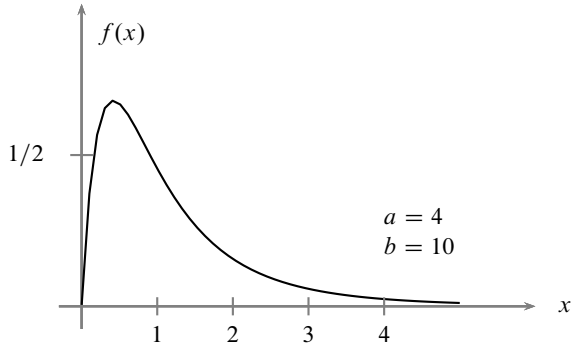


FIGURA 1.39. Función de densidad $F(a, b)$.

La distribución F aparece como resultado de la siguiente operación entre variables aleatorias.

PROPOSICIÓN 1.104. *Sean $X \sim \chi^2(a)$ y $Y \sim \chi^2(b)$ variables aleatorias independientes. Entonces*

$$\frac{X/a}{Y/b} \sim \mathrm{F}(a, b).$$

Con esto terminamos con una revisión elemental de algunas distribuciones de probabilidad continuas. Recuerde el lector que existen muchas mas distribuciones de este tipo, algunas más conocidas que otras, pero todas ellas útiles como modelos probabilísticos en las muy diversas áreas de aplicación de la probabilidad. En la segunda parte del texto usaremos algunas de estas distribuciones al plantear y resolver ciertos problemas dentro de la estadística matemática.

9. Dos teoremas límite

En esta sección enunciaremos sin demostración dos de los teoremas límite más importantes en la probabilidad: la ley de los grandes números y el teorema central del límite. Mostraremos algunos ejemplos de aplicación y en la segunda parte del libro haremos uso de ellos. Los enunciados de estos resultados mencionan el concepto de independencia de varias variables aleatorias, el cual será revisado más adelante en la página 107. Dicho concepto está estrechamente ligado al concepto de independencia de eventos estudiado antes.

Ley de los grandes números

Este interesante resultado establece que bajo ciertas condiciones, el promedio de variables aleatorias converge a una constante cuando el número de sumandos crece a infinito.

TEOREMA 1.105 (Ley de los grandes números). *Sea $X_1, X_2, \ldots$ una sucesión infinita de variables aleatorias independientes e idénticamente distribuidas con media finita μ. Entonces*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \longrightarrow \mu.$$

Así, sin importar la distribución de las variables aleatorias, el promedio converge a la media μ conforme n tiende a infinito. Damos a continuación dos ejemplos de aplicación de este interesante resultado.

EJEMPLO 1.106 (Probabilidad frecuentista). Considere un experimento aleatorio cualquiera y sea A un evento con probabilidad p. Suponga que se efectúan realizaciones independientes del experimento y se observa en cada ensayo la ocurrencia o no ocurrencia del evento A. Para cada entero $k \geq 1$ defina la variable aleatoria

$$X_k = \begin{cases} 1 & \text{si ocurre el evento } A \text{ en el } k\text{-ésimo ensayo,} \\ 0 & \text{si no ocurre el evento } A \text{ en el } k\text{-ésimo ensayo.} \end{cases}$$

Entonces las variables $X_1, X_2, \ldots$ son independientes cada una con distribución $\text{Ber}(p)$, en donde p es la probabilidad del evento A. Por lo tanto $E(X_k) = p$ y $\text{Var}(X_k) = p(1-p)$. La ley de los grandes números asegura que la fracción de ensayos en los que se observa el evento A converge a la constante desconocida p cuando el número de ensayos crece a infinito, es decir,

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} X_i = \lim_{n\to\infty} \frac{n_A}{n} = p.$$

Esta es la definición frecuentista de la probabilidad que hemos estudiado en la página 9 y hemos ahora corroborado su validez con ayuda de la ley de los grandes números.

EJEMPLO 1.107 (Simulación). En este ejemplo se simularon 200 lanzamientos independientes de una moneda equilibrada, es decir, se generaron 200 valores de una v.a. con distribución Ber(p), con $p = 1/2$. En la Figura 1.40 se muestra el comportamiento asintótico de los promedios $(X_1 + \cdots + X_n)/n$ conforme n crece. Los puntos graficados fueron unidos por una línea continua para una mejor visualización del comportamiento inicial oscilante y su eventual estabilización en el valor $1/2$. En este ejemplo hemos utilizado la distribución Bernoulli pero cualquier otra distribución puede ser usada para observar el comportamiento asintótico a la media. En la última parte de este capítulo estudiaremos la forma de simular valores de una variable aleatoria en computadora.

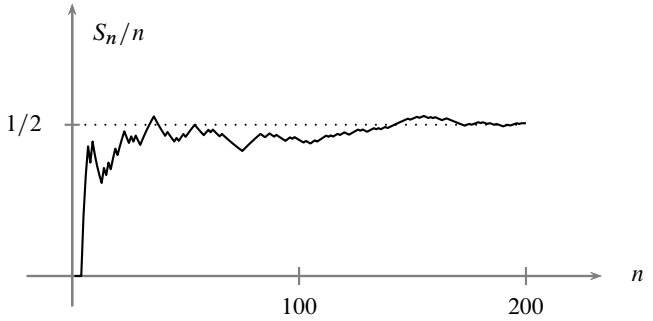


FIGURA 1.40. Ley de los grandes números.

Teorema central del límite

Este teorema es muy importante y tiene una amplia gama de aplicaciones. En la segunda parte del texto veremos varios ejemplos y situaciones que se presentan en el estudio de la estadística matemática, en donde se usa este resultado para aproximar y simplificar el cálculo de algunas probabilidades.

TEOREMA 1.108 (Teorema central del límite). *Sea $X_1, X_2, \ldots$ una sucesión infinita de variables aleatorias independientes e idénticamente distribuidas, con media μ y varianza finita σ^2. Entonces la función de distribución de la variable aleatoria*

$$Z_n = \frac{(X_1 + \cdots + X_n) - n\mu}{\sqrt{n\sigma^2}}$$

tiende a la función de distribución normal estándar cuando n tiende a infinito.

En términos matemáticos este resultado establece que para cualquier número real x,

$$\lim_{n\to\infty} F_{Z_n}(x) = F_Z(x),$$

sin importar la distribución de las variables $X_1, X_2, \ldots$, así es que éstas pueden tener distribución Bernoulli, binomial, exponencial, gama, etc., en general pueden ser discretas o continuas, y siempre se tendrá que Z_n tiene una distribución aproximada normal estándar. Esto nos permitirá aproximar probabilidades de eventos que involucren sumas de variables aleatorias en términos de probabilidades de la distribución normal estándar. Observe que dividiendo el numerador y denominador entre n, y definiendo $\bar{X} = (X_1 + \cdots + X_n)/n$, la variable Z_n puede escribirse de la siguiente forma

$$Z_n = \frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}}.$$

EJEMPLO 1.109. Se lanza una dado repetidas veces y sean $X_1, X_2, \ldots$ los resultados de estos lanzamientos. Es razonable suponer que estas variables aleatorias son independientes y con idéntica distribución uniforme en el conjunto $\{1, 2, 3, 4, 5, 6\}$. En particular, la esperanza es $\mu = 3.5$ y la varianza es $\sigma^2 = 2.91\overline{6}$. Por la ley de los grandes números, sabemos que el promedio parcial $\bar{X} = (X_1 + \cdots + X_n)/n$ se aproxima a la media 3.5 conforme n crece. ¿Cuántas veces debe lanzarse el dado de tal forma que $\bar{X}$ se encuentre entre 3 y 4 con una probabilidad de 0.99?
Solución. Se busca el valor de n tal que

$$P(3 \leq \bar{X} \leq 4) = 0.99\,.$$

Restando en cada lado de las desigualdades la media μ y dividiendo entre $\sqrt{\sigma^2/n}$, la igualdad anterior es equivalente a la ecuación

$$P(\frac{3 - 3.5}{\sqrt{\sigma^2/n}} \leq \frac{\bar{X} - 3.5}{\sqrt{\sigma^2/n}} \leq \frac{4 - 3.5}{\sqrt{\sigma^2/n}}) = 0.99\,.$$

Por el teorema central del límite, la probabilidad indicada es aproximadamente igual a $P(-0.5/\sqrt{\sigma^2/n} \leq Z \leq 0.5/\sqrt{\sigma^2/n})$, en donde Z es una variable aleatoria con distribución normal estándar. Es decir, tenemos ahora la ecuación de aproximación

$$\Phi(\frac{0.5}{\sqrt{\sigma^2/n}}) - \Phi(\frac{-0.5}{\sqrt{\sigma^2/n}}) = 0.99\,.$$

De tablas de la distribución normal estándar puede verificarse que el valor de x tal que $\Phi(x) - \Phi(-x) = 0.99$ es $x = 2.58$. De este modo se tiene que $0.5/\sqrt{\sigma^2/n} = 2.58$, de donde se obtiene $n = 226.5$.

EJEMPLO 1.110. Se desea diseñar un estacionamiento de coches para un conjunto de 200 departamentos que se encuentran en construcción. Suponga que para cada departamento, el número de automóviles será de 0, 1 o 2, con probabilidades 0.1, 0.6 y 0.3, respectivamente. Se desea que con una certeza del 95 % haya espacio disponible para todos los coches cuando los departamentos se vendan. ¿Cuántos espacios de estacionamiento deben construirse?
Solución. Sean $X_1, \ldots, X_{200}$ las variables aleatorias que denotan el número de automóviles que poseen los futuros dueños de los departamentos. Podemos suponer

que estas variables aleatorias discretas son independientes unas de otras y todas ellas tienen la misma distribución de probabilidad: $P(X = 0) = 0.1$, $P(X = 1) = 0.6$, y $P(X = 2) = 0.3$. De esta forma la variable aleatoria suma $X_1 + \cdots + X_{200}$ denota el total de automóviles que habrá en el complejo de departamentos. Se desconoce la distribución de esta variable aleatoria, sin embargo se desea encontrar el valor de n tal que $P(X_1 + \cdots + X_{200} \leq n) = 0.95$. Haremos uso del teorema central del límite para resolver este problema, y para ello se necesita calcular la esperanza y varianza de X. Puede comprobarse que $E(X) = 1.2$ y $\text{Var}(X) = 0.36$, cantidades que denotaremos por μ y σ^2 respectivamente. La ecuación planteada es entonces

$$P(X_1 + \cdots + X_{200} \leq n) = 0.95\,,$$

en donde la incógnita es el valor de n. Restando en ambos lados de la desigualdad 200μ y dividiendo entre $\sqrt{200\sigma^2}$, la ecuación anterior es equivalente a

$$P\left(\frac{X_1 + \cdots + X_{200} - 200\mu}{\sqrt{200\sigma^2}} \leq \frac{n - 200\mu}{\sqrt{200\sigma^2}}\right) = 0.95\,. \tag{6}$$

Por el teorema central del límite, la probabilidad indicada es aproximada a $\Phi((n - 200\mu)/\sqrt{200\sigma^2})$. De este modo tenemos ahora la ecuación

$$\Phi\left(\frac{n - 200\mu}{\sqrt{200\sigma^2}}\right) = 0.95\,.$$

Observe que la variable aleatoria que aparece enla ecuación (6) y cuya distribución de probabilidad se desconoce y en general es difícil conocer, se ha aproximado por una variable aleatoria normal estándar, y allí radica la utilidad del teorema central del límite. De la tabla de la distribución normal podemos ahora verificar que el valor de x tal que $\Phi(x) = 0.95$ es $x = 1.65$. De este modo se llega a la igualdad $(n - 200\mu)/\sqrt{200\sigma^2} = 1.65$, de donde se obtiene que $n = 253.99$. Es decir, el tamaño del estacionamiento debe ser de aproximadamente 254 lugares.

10. Simulación

Esta sección contiene una breve introducción al tema de simulación de variables aleatorias. En los actuales paquetes computacionales existen funciones predefinidas que proporcionan de manera inmediata valores al azar de las distintas distribuciones de probabilidad. Por ejemplo, la función `random()` proporciona un valor aleatorio con distribución uniforme dentro del intervalo $[0, 1)$ en el lenguaje Python:

```
>>> import random
>>> random.random()
0.6386575215538782
```

Existen expresiones equivalentes para otros paquetes de cómputo con muy diversas opciones para ésta y otras distribuciones de probabilidad de uso común. No obstante que contamos con una inmediata disponibilidad de estas herramientas, el objetivo de

esta sección es exponer los fundamentos matemáticos que subyacen en algunas de las implementaciones de los mecanismos para generar números aleatorios. En varios de estos procedimientos se tiene como punto de partida el saber generar números dentro del intervalo unitario $[0, 1]$ con distribución uniforme. Hay dos aspectos importantes por resaltar. Primero, la mayoría de los métodos utilizados para generar números aleatorios no son completamente aleatorios, de modo que los resultados que se obtienen se denominan seudoaleatorios; en el siguiente párrafo daremos un ejemplo de ello. Segundo, debido a que las computadoras tienen una precisión limitada para representar a los números reales, no todos los números del intervalo $[0, 1]$, por ejemplo, son susceptibles de ser generados como números aleatorios, sólo un número finito de ellos pueden ser escogidos. Si para algún entero m la fracción $1/m$ es el número positivo más pequeño que una computadora puede representar, entonces los únicos números en el intervalo $[0, 1]$ que pueden representarse son $0, 1/m, 2/m, 3/m, \ldots, m/m$. Teniendo estas consieraciones en mente, revisaremos a continuación un mecanismo particular que puede usarse para generar números seudoaleatorios enteros.

Simulación de v.a.s discretas

El siguiente procedimiento puede ser usado para generar valores seudoaleatorios dentro del conjunto $\{0, 1, 2, \ldots, m\}$, en donde m un entero positivo fijado de antemano. Sean a y b dos enteros fijos, y sea X_0 cualquier valor dentro del conjunto indicado. Para cada $n \geq 0$ se calcula de manera iterada

$$X_{n+1} := aX_n + b \quad \text{módulo } (m + 1). \tag{7}$$

Por ejemplo, si se toma $m = 100$, $a = 5$, $b = 13$ y $X_0 = 23$ se obtiene la sucesión de números enteros:

$$23, 27, 47, 46, 41, 16, 93, 74, 80, 9, 58, 0, 13, 78, 100, 8, 53, 76, 90, 59, \ldots$$

El código Python para obtener esta sucesión usando la ecuación (7) es el siguiente:

```
>>> m=100
>>> a,b=5,13
>>> n=20 # Numero de iteraciones
>>> x=23 # Valor inicial
>>> for j in range (1,n+1):
>>>         print x,
>>>         x=(a*x+b)%(m+1) # El operador % calcula modulo
```

Es importante observar que la sucesión numérica que se obtiene a partir de (7) es periódica, pues siendo el algoritmo completamente determinista, en cuanto un valor generado aparece por segunda vez, la cadena de enteros previamente obtenida se repite una infinidad de veces. Es conveniente, por tanto, escoger un valor grande para m a fin de que este fenómeno de repetición no aparezca demasiado pronto. Por ejemplo, si se toma $m = 10$, $a = 9$, $b = 9$ y $X_0 = 1$ se obtiene la sucesión:

$$1, 7, 6, 8, 4, 1, 7, 6, 8, 4, 1, \ldots$$

Por otro lado, si se toma $m = 10$, $a = 1$, $b = 1$ y $X_0 = 0$ se obtiene la bonita sucesión:

$$0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 0, \ldots$$

la cual no presenta ninguna característica de ser aleatoria. Así pues, deben escogerse adecuadamente los parámetros de la ecuación (7) para obtener resultados medianamente aceptables. Si X es un valor seudoaleatorio generado mediante el mecanismo descrito y si m es muy grande, entonces podemos considerar que $U = X/m$ es un valor seudoaleatorio del intervalo $[0, 1]$. No tomando en cuenta la aparición de los valores 0 ó m, pueden también generarse números seudoaleatorios en los intervalos $[0, 1)$, $(0, 1]$ y $(0, 1)$.

Supongamos entonces que contamos con un mecanismo para generar números al azar dentro del intervalo $[0, 1]$ con distribución uniforme, sea este mecanismo implementado por nosotros mismos o usando alguna función de un paquete computacional. A partir de esta herramienta veremos la forma en la que se pueden generar valores al azar de una variable aleatoria con cualquier distribución discreta. Veamos primero el caso de la distribución Bernoulli.

Distribución Bernoulli. A partir de un valor de U con distribución unif$[0, 1]$, puede crearse un valor de X con distribución Ber(p) de la siguiente forma:

$$X = \begin{cases} 1 & \text{si } U \in [0, p], \\ 0 & \text{si } U \in (p, 1], \end{cases}$$

en donde estamos suponiendo que el valor del parámetro p es conocido. Véase la Figura 1.41.

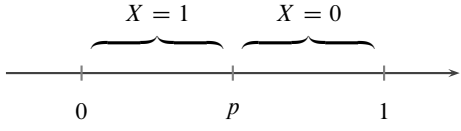


FIGURA 1.41

De esta forma efectivamente se verifica que X toma los valores 0 y 1 con las probabilidades requeridas:

$$P(X = 1) = P(U \in [0, p]) = p,$$
$$P(X = 0) = P(U \in (p, 1]) = 1 - p.$$

Distribución binomial. Conociendo una forma de obtener observaciones de una variable aleatoria Bernoulli, podemos generar observaciones de una variable aleatoria X con distribución binomial sumando n observaciones independientes Bernoulli $X_1, X_2, \ldots, X_n$, es decir,

$$X = \sum_{i=1}^{n} X_i.$$

Distribuciones geométrica y binomial negativa. Dado que las distribuciones geo(p) y binneg(r, p) se definen a partir de una sucesión de variables independientes Bernoulli, a saber, como el número de fracasos (o ceros) antes de obtener el r-ésimo éxito (o uno), entonces si n es el número de ensayos tal que $X_1 + \cdots + X_n = r$ por primera vez, un valor para X con distribución binneg(r, p) se calcula de la siguiente forma:

$$X = \sum_{i=1}^{n} X_i - r.$$

El caso $r = 1$ corresponde a la distribución geo(p).

Distribución Poisson. Explicaremos a continuación un método para generar valores al azar de la distribución Poisson(λ) a partir de la distribución exponencial. Sea $T_1, T_2, \ldots$ una sucesión de variables aleatorias independientes con idéntica distribución exp(λ). Defina la variable aleatoria discreta

$$N = \begin{cases} 0 & \text{si } T_1 > 1, \\ k & \text{si } T_1 + \cdots + T_k \leq 1 < T_1 + \cdots + T_{k+1}. \end{cases}$$

Es un buen ejercicio comprobar que la variable N tiene efectivamente una distribución Poisson(λ). Así, para generar valores de la distribución Poisson, este procedimiento sugiere que se deben generar tantos valores al azar de la distribución exponencial hasta que la suma de todos ellos exceda el valor 1 por primera vez. El valor asignado a N es el número de valores exponenciales generados para que se cumpla la condición indicada menos uno. Por supuesto, para aplicar este procedimiento se necesita saber generar valores al azar de la distribución exponencial pero esto lo estudiaremos en la siguiente sección y veremos que tal distribución continua es una de las más sencillas de simular.

Ahora extenderemos el método explicado para obtener valores de la distribución Bernoulli (Figura 1.41) al caso de cualquier variable aleatoria discreta.

Cualquier distribución discreta. Suponga que X es una variable con posibles valores y probabilidades como se indica en la siguiente tabla:

x	x_1	x_2	x_3	$\cdots$
$P(X = x)$	p_1	p_2	p_3	$\cdots$

A partir de un valor de la variable U con distribución unif[0, 1], puede generarse un valor de la variable X de la siguiente forma:

$$X = \begin{cases} x_1 & \text{si } 0 \leq U \leq p_1, \\ x_2 & \text{si } p_1 < U \leq p_1 + p_2, \\ x_3 & \text{si } p_1 + p_2 < U \leq p_1 + p_2 + p_3, \\ \vdots & \vdots \\ x_n & \text{si } p_1 + \cdots + p_{n-1} < U \leq p_1 + \cdots + p_n, \\ \vdots & \vdots \end{cases}$$

Efectivamente se verifica que la v.a. X así definida tiene la distribución indicada, pues para cada $n \geq 1$,

$$P(X = x_n) = P(p_1 + \cdots + p_{n-1} < U \leq p_1 + \cdots + p_n) = p_n.$$

En la práctica este mecanismo general funciona perfectamente bien para v.a.s discretas que toman un número no muy grande de valores, pero la situación podría ser un tanto distinta en el caso cuando la variable toma una infinidad de valores, la dificultad radicará nuevamente en la imprecisión para representar probabilidades muy pequeñas.

Por otro lado, debemos mencionar también que si contamos con un mecanismo para generar números aleatorios de una variable aleatoria, entonces también se pueden generar números al azar de una transformación de esta variable aleatoria, es decir, si X una variable aleatoria y g es una función tal que $g(X)$ es otra variable, por ejemplo, $g(X) = X^2$, o cualquier otra expresión analítica, entonces si x es un valor al azar de X, entonces $g(x)$ representa un valor al azar de la variable $Y = g(X)$. Este mecanismo es bastante útil y se aplica para variables aleatorias tanto discretas como continuas.

Veremos a continuación algunos mecanismos de simulación de variables aleatorias continuas.

Simulación de v.a.s continuas

Consideremos primero el caso de la distribución uniforme continua.

Distribución uniforme continua. Si U es una variable con distribución unif$[0, 1]$, entonces la variable aleatoria X definida de la siguiente manera tiene distribución unif$[a, b]$.

$$X = a + (b - a)U.$$

Método de la transformación inversa. El mecanismo explicado para la distribución uniforme continua puede generalizarse usando el siguiente resultado que se conoce como el método de la transformación inversa.

PROPOSICIÓN 1.111 (Método de la transformación inversa). *Sea X una variable aleatoria con función de distribución $F(x)$ continua, estrictamente creciente y para la cual se conoce una expresión para la función inversa $F^{-1}(u)$. Si U tiene distribución* unif$[0, 1]$*, entonces la variable $F^{-1}(U)$ tiene la misma distribución que X.*

DEMOSTRACIÓN. Sea Y la variable aleatoria $F^{-1}(U)$. Entonces para cualquier valor y de Y,

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(F^{-1}(U) \leq y) \\ &= P(U \leq F(y)) \\ &= F_U(F(y)) \\ &= F(y). \end{aligned}$$

La última igualdad se debe a que U tiene distribución uniforme en el intervalo unitario. ■

Así, generando valores para U y evaluando en la expresión $F^{-1}(U)$ se obtienen valores para X. En la Figura 1.42 se muestra gráficamente este procedimiento.

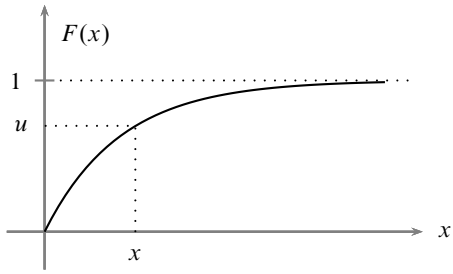


FIGURA 1.42

Por ejemplo, para el caso cuando X tiene distribución unif$[a, b]$, su función de distribución es

$$F(x) = \frac{x-a}{b-a}, \quad a \leq x \leq b,$$

que es estrictamente creciente en dicho intervalo y cuya función inversa es

$$F^{-1}(u) = a + (b-a)u, \quad 0 \leq u \leq 1.$$

Este resultado había sido mencionado al inicio de esta sección.

Distribución exponencial. Otro caso para el cual se puede aplicar el método anterior es para la distribución exponencial: sea X una variable con distribución exp(λ). Entonces su función de distribución está dada por

$$F(x) = 1 - e^{-\lambda x}, \quad x > 0.$$

La función inversa es

$$F^{-1}(u) = -\frac{1}{\lambda} \ln(1-u), \quad 0 \leq u < 1.$$

Por lo tanto, si u es un valor de la distribución unif$[0, 1)$, entonces $F^{-1}(u)$ es un valor de la distribución exp(λ). El código Python para obtener una observación exponencial es

```
>>> import random, math
>>> u=random.random()
>>> lamda=2.0
>>> -(1/lamda)*math.log(1-u)
0.23962856605708488
```

Distribución gama. Cuando el parámetro n es un número natural, usando la Proposición 1.91 en la página 63, se pueden generar valores al azar de la distribución gama(n, λ) sumando n valores independientes de la distribución exp(λ).

Distribución beta. La Proposición 1.92 que hemos mencionado sin demostración en la página 64 establece que si X y Y son independientes con distribución gama(a, λ) y gama(b, λ), respectivamente, entonces la variable aleatoria $X/(X+Y)$ tiene distribución beta(a, b). Así, suponiendo que podemos generar valores al azar x y y de las distribuciones gama mencionadas, entonces el cociente $x/(x+y)$ será un valor al azar de la distribución beta(a, b).

Distribución Weibull. Otro ejemplo adicional para el cual se aplica el método de la transformación inversa es para la distribución Weibull(λ, α). Habíamos mencionado que la correspondiente función de distribución es

$$F(x) = 1 - e^{-(\lambda x)^{\alpha}}, \quad x > 0.$$

Por lo tanto la función inversa es

$$F^{-1}(u) = -\frac{1}{\lambda}\,[\ln(1-u)\,]^{1/\alpha}, \quad 0 \leq u < 1.$$

Así, nuevamente, si u es un valor de la distribución unif$[0, 1)$, entonces $F^{-1}(u)$ es un valor de la distribución Weibull(λ, α).

Un ejemplo importante para el cual no puede aplicarse el método de la función inversa es la distribución normal, para esta distribución no existe una expresión analítica para la función de distribución y por lo tanto tampoco existe tal expresión para su inversa. Sin embargo, el siguiente método que expondremos de manera general, permite en particular obtener observaciones de tan importante variable aleatoria.

Método de aceptación o rechazo. El siguiente procedimiento permite generar observaciones de una variable aleatoria X con cualquier función de densidad $f(x)$ cuando se cumplen ciertas condiciones. Se necesita contar con otra variable aleatoria Y para la cual se conozca un mecanismo para generar números al azar.

PROPOSICIÓN 1.112 (Método de aceptación o rechazo). *Sean X y Y variables aleatorias con función de densidad $f(x)$ y $g(y)$ respectivamente. Suponga que existe una constante $c > 1$ tal que para todo x real,*

$$f(x) \leq c\, g(x).$$

Sea U una variable aleatoria con distribución unif$[0, 1]$ *e independiente de la variable Y. Entonces, condicionada al cumplimiento de la desigualdad que aparece abajo, la variable aleatoria Y tiene la misma distribución que X.*

$$U \leq \frac{f(Y)}{c\, g(Y)}. \tag{8}$$

De esta manera, el método de aceptación y rechazo para generar valores de la variable X consiste en los siguientes pasos:

a) Se genera un valor de U con distribución unif$[0, 1]$.
b) Se genera un valor de Y independiente de U.
c) Si se cumple la desigualdad (8) entonces se acepta el valor de Y como valor de X, de lo contrario se vuelven a generar valores para U y para Y, y se repite el procedimiento de aceptación o rechazo hasta obtener el número necesario de valores aceptados.

Daremos a continuación los detalles de la prueba de la Proposición 1.112, en donde se utilizan algunos conceptos y herramientas un tanto más avanzadas que las que hemos estudiado hasta ahora, El lector, por tanto, no debe preocuparse en caso de que no entender el origen y desarrollo de estos cálculos. En esta demostración se muestra que efectivamente los valores aceptados de Y tienen la misma distribución que X.

DEMOSTRACIÓN. La probabilidad de obtener una aceptación en cada intento es, por el teorema de probabilidad total en su versión continua,

$$\begin{aligned} P(U \leq \frac{f(Y)}{c\, g(Y)}) &= \int_{-\infty}^{\infty} P(U \leq \frac{f(y)}{c\, g(y)})\, g(y)\, dy \\ &= \int_{-\infty}^{\infty} \frac{f(y)}{c\, g(y)}\, g(y)\, dy \\ &= \frac{1}{c} \int_{-\infty}^{\infty} f(y)\, dy \\ &= \frac{1}{c}. \end{aligned}$$

De modo que la función de probabilidad del número de ensayos hasta obtener un éxito es $f(x) = (1-p)^{x-1}\, p$, para $x = 1, 2, \ldots$, en donde $p = 1/c$. La media de esta distribución es $c = 1/p$, así es que mientras más pequeño sea el valor de c más frecuentemente se tendrán aceptaciones. La distribución para el número de ensayos hasta obtener un éxito que hemos mencionado corresponde a la distribución de una v.a. geométrica (así como la hemos definido en este texto) mas uno. Ahora veamos que la distribución condicional de Y, dado que ha tomado un valor aceptado, es la misma que la distribución de X. Por independencia, la función de densidad conjunta de U y Y es

$f(u, y) = g(y)$. Por lo tanto,

$$\begin{aligned} P(Y \leq x \,|\, U \leq \frac{f(Y)}{c\, g(Y)}) &= \frac{P(Y \leq x, U \leq \frac{f(Y)}{c\, g(Y)})}{P(U \leq \frac{f(Y)}{c\, g(Y)})} \\ &= c\, P(Y \leq x, U \leq \frac{f(Y)}{c\, g(Y)}) \\ &= c \int_{-\infty}^{x} \int_{0}^{f(y)/cg(y)} g(y)\, du\, dy \\ &= c \int_{-\infty}^{x} f(y)\, dy \\ &= P(X \leq x). \end{aligned}$$

■

Distribución normal. Como ejemplo de aplicación del método de aceptación o rechazo, explicaremos la forma de obtener observaciones de una variable aleatoria X con distribución normal estándar. Puede demostrarse que la función de densidad de la variable $|X|$ es

$$f(x) = \sqrt{\frac{2}{\pi}}\, e^{-x^2/2}, \quad x \geq 0.$$

Por otro lado, considere la v.a. Y con distribución $\exp(\lambda)$, con $\lambda = 1$, su función de densidad es por lo tanto

$$g(y) = e^{-\lambda y}, \quad y > 0.$$

Se comprueba que existe una constante c tal que $f(x)/g(x) \leq c$, en efecto,

$$\frac{f(x)}{g(x)} = \sqrt{\frac{2}{\pi}}\, e^{-\frac{x^2}{2}+x} = \sqrt{\frac{2e}{\pi}}\, e^{-\frac{1}{2}(x-1)^2} \leq \sqrt{\frac{2e}{\pi}}.$$

Esto demuestra que puede tomarse $c = \sqrt{2e/\pi}$. De esta forma el procedimiento para generar valores de $|X|$ es el siguiente: se genera un valor u con distribución unif$[0, 1]$, se genera un valor y independiente de u, cuando se cumple la desigualdad de abajo se acepta a y como valor de $|X|$, de lo contrario se repite este procedimiento hasta obtener el número de valores aceptados necesario para $|X|$.

$$e^{-\frac{1}{2}(y-1)^2} \geq u.$$

Sabiendo la forma de generar valores para $|X|$, podemos ahora generar valores para X a través de una v.a. V con distribución $\text{Ber}(p)$ de parámetro $p = 1/2$, e independiente de las variables anteriores, de la forma siguiente:

$$X = \begin{cases} +|X| & \text{si } V = 0, \\ -|X| & \text{si } V = 1. \end{cases}$$

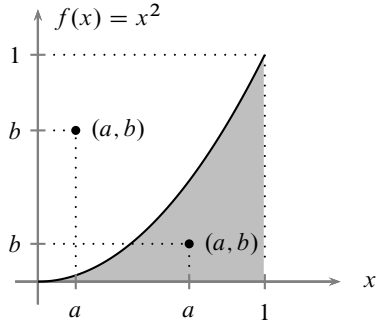
De esta manera X tiene distribución normal estándar. Finalmente, al evaluar en la expresión $\mu + \sigma X$ se obtienen valores de la distribución $\text{N}(\mu, \sigma^2)$.

Distribución ji-cuadrada. Usando las Proposiciones 1.99 y 1.100 en la página 70, a partir de observaciones de la distribución normal estándar pueden obtenerse simulaciones de la distribución ji-cuadrada.

Distribución t. Usando la Proposición 1.102 en la página 71, a partir de observaciones de las distribuciones normal estándar y ji-cuadrada, pueden obtenerse observaciones de la distribución t.

Distribución F. Usando la Proposición 1.104 en la página 72 se pueden obtener simulaciones de la distribución F a partir de observaciones de la distribución ji-cuadrada.

SIMULACIÓN 1.113. En este proyecto se busca aproximar la integral de la función $f(x) = x^2$ en el intervalo $[0, 1]$ a través de valores generados al azar de la distribución unif$(0, 1)$.



$$\int_0^1 x^2\,dx = \frac{1}{3}$$

FIGURA 1.43

Para lograr esta aproximación se pueden generar parejas de valores (a, b) en donde a y b son valores al azar de la distribución unif$(0, 1)$. Cuando a y b son tales que $b \leq a^2$, el punto (a, b) se encuentra en la región sombreada de la Figura 1.43 y se cuenta como un éxito, en caso contrario se considera como un fracaso. Si se lleva a cabo este experimento repetidas veces y se calcula el cociente del número de éxitos entre el número de ensayos realizados, entonces se obtiene una aproximación de la integral, que es $1/3$.

SIMULACIÓN 1.114. En este experimento se pide aproximar el valor de π de la siguiente forma: se generan parejas (a, b) en donde nuevamente a y b son valores al azar de la distribución unif$(0, 1)$. Si ocurre que $b \leq \sqrt{1 - a^2}$, entonces el punto (a, b) pertenece a la región sombreada de la Figura 1.44 y se considera un éxito, en caso contrario se considera un fracaso. Así, al repetir varias veces este procedimiento el cociente del número de éxitos entre el número de ensayos será una aproximación del área sombreada de la Figura 1.44. Siendo esta área una cuarta parte del círculo unitario, el área es $\pi/4$.

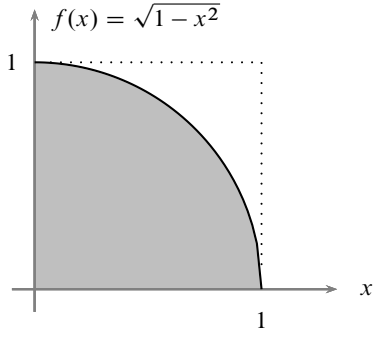


FIGURA 1.44

SIMULACIÓN 1.115. Asigne un valor para $\lambda > 0$ y genere al azar 50 valores de la distribución $\exp(\lambda)$ usando el método de la transformación inversa. Compruebe la ley de los grandes números, es decir, verifique que el promedio $(x_1 + \cdots + x_{50})/50$ es cercano a $1/\lambda$.

Capítulo 2

Estadística

Originalmente cualquier colección de datos acerca de la población o economía de un estado era del interés exclusivo del estado, por ello es que a las primeras colecciones de datos se les llamara estadísticas, como una derivación de la palabra estado.

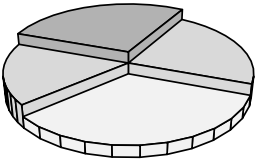
Se tienen indicios de que los grandes imperios de la antigüedad realizaron registros de algunas de sus actividades económicas y demográficas. La civilización Babilónica (5000 a.C.), el Egipto de la Faraones (siglo VI a.C.), La India (siglo IV a.C.), Los Griegos, Los Romanos y Los Incas, por ejemplo, llevaron a cabo en algún momento de su desarrollo, listas de personas y familias, tierras, bienes, soldados, armas, tributos de guerra, etc. Desde tiempos antiguos las grandes civilizaciones se dieron cuenta que contar con tal información les permitía tener mas control y una mejor administración de todo lo relacionado con su imperio.

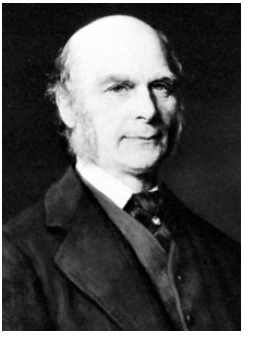
En Europa, los primeros registros de una recolección sistemática de datos de una población surgen durante el Renacimiento en las ciudades italianas de Venecia y Florencia, y después tales prácticas se extienden a otras ciudades europeas. Para mediados del siglo XVI, varios gobiernos europeos, a través de las parroquias, llevaban registros de nacimientos, matrimonios y decesos. De particular importancia en aquellos años eran los registros de decesos pues las tasas de mortalidad eran con frecuencia altas debido a las recurrentes epidemias y plagas que diezmaban de manera importante a la población. Esta recolección de datos por parte de los gobiernos siguió desarrollándose durante los siglos XVII y XVIII, en donde los esfuerzos y recursos se orientaron principalmente a obtener censos, es decir, información de una población completa. A finales del siglo XIX y teniendo a Francis Galton (1822-1911) como precursor y a personajes como Adolphe Quételet (1796-1874) y Florence Nightingale (1820-1910), por nombrar sólo algunos, surgen las primeras ideas para inferir información de toda una población a partir de datos estadísticos parciales. Los métodos estadísticos y su formalización como teoría matemática se desarrollaron en los inicios del siglo XX gracias a los trabajos de Karl Pearson (1857-1936), Ronald A. Fisher (1890-1962), Egon Sharpe Pearson (1895-1980) hijo de Karl Pearson, y Jerzy Neyman (1894-1981), entre muchos otros matemáticos y estadísticos importantes.

Hoy en día los métodos de la estadística son usados con mucha frecuencia y con bastante provecho en muy diversas áreas del conocimiento humano. Sus métodos y

conclusiones son usados como respaldo científico para rechazar o aceptar afirmaciones, para determinar tendencias o comportamientos y también para tomar decisiones importantes tanto a nivel personal como organizacional, incluyendo las importantes decisiones gubernamentales. Por otro lado, de manera cotidiana, todo ciudadano se encuentra expuesto a información estadística a través de los periódicos, la radio, la televisión e internet; en donde es frecuente encontrar información resumida en forma de gráficas y tablas, así como la presentación del comportamiento cualitativo o cuantitativo de ciertas variables de interés. Esta información requiere de una correcta interpretación y una razonada evaluación, y ése justamente es el tema de la segunda parte de nuestro curso: comprender algunos conceptos y técnicas elementales de la estadística matemática.

Florence Nightingale
(Inglaterra 1820–1910)





Francis Galton
(Inglaterra 1822–1911)

1. Introducción

Supongamos que tenemos una población de interés, esto es, un conjunto arbitrario de personas, mediciones u objetos cualesquiera, y deseamos conocer cierta información de esta población. Debido a la imposibilidad o no conveniencia de tener información de todos y cada uno de los elementos de la población, tomamos un pequeño subconjunto de ellos, al cual llamaremos muestra. Con base en esta muestra trataremos de inferir la información de la población en su totalidad. El llevar a cabo tal ejercicio de inferencia a partir de información parcial traerá consigo el riesgo de cometer errores. A pesar de ello, es preferible alguna información o afirmación con una cierta confianza cercana al 100 % en su veracidad que ninguna información con total certidumbre. El objetivo también es buscar cuantificar el posible error en nuestras afirmaciones y procurar que los errores tengan poca probabilidad de ocurrir.

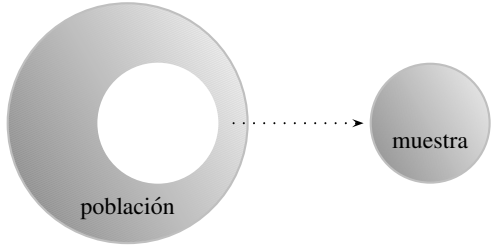


FIGURA 2.1

Estadística descriptiva e inferencial

Daremos ahora una definición de la estadística como rama de estudio y después daremos una primera clasificación en dos de sus quehaceres generales. La estadística es la ciencia que se encarga de recolectar, organizar, resumir y analizar datos para después obtener conclusiones a partir de ellos. De manera general, la estadística puede ser dividida en dos grandes áreas: estadística inferencial y estadística descriptiva. La estadística descriptiva es una colección de métodos para la organización, resumen y presentación de datos. La estadística inferencial, en cambio, consiste de un conjunto de técnicas para obtener, con determinado grado de confianza, información de una población con base en la información de una muestra.

2. Variables y tipos de datos

Una variable es una característica de interés de un elemento en una población en estudio. Por ejemplo, si la población consta de personas, los siguientes son ejemplos de variables que podrían ser de interés: edad, peso, sexo, estatura, estado civil, ingreso económico, preferencia electoral, etc. Veremos a continuación algunos criterios bajo los cuales las variables pueden ser clasificadas. Las variables pueden ser cuantitativas, cuando se realiza una medición y el resultado es un número, o pueden ser cualitativas, cuando solamente registran una cualidad o atributo del objeto o persona en estudio. La edad, el peso y la estatura son ejemplos de variables cuantitativas en una población de personas, mientras que el sexo y el estado civil son variables cualitativas.

$$\text{Variable} \begin{cases} \text{cualitativa} \\ \text{cuantitativa} \end{cases}$$

Por otro lado, de acuerdo al número total de sus posibles valores, una variable cuantitativa puede ser clasificada como discreta cuando sólo puede tomar un número discreto (es decir, finito o numerable) de valores, o continua, cuando puede tomar cualquier valor dentro de un intervalo (a, b) de la recta real.

$$\text{Variable cuantitativa} \begin{cases} \text{discreta} \\ \text{continua} \end{cases}$$

De acuerdo con la posible relación que pudieran guardar los valores de una variable, se cuenta por lo menos con cuatro escalas de medición. Las variables cualitativas pueden ser clasificadas de acuerdo a dos escalas: escala nominal o escala ordinal,

$$\text{Variable cualitativa} \begin{cases} \text{Escala nominal} \\ \text{Escala ordinal} \end{cases}$$

Mientras que las variables cuantitativas pueden clasificarse mediante la escala de intervalo o la escala de razón.

$$\text{Variable cuantitativa} \begin{cases} \text{Escala de intervalo} \\ \text{Escala de razón} \end{cases}$$

Escala nominal

Se dice que una variable es de tipo nominal cuando sus posibles valores no tienen ninguna relación de orden o magnitud entre ellos. Los valores de este tipo de variables son etiquetas sin un orden, por ejemplo, si estamos estudiando una población humana, a la variable sexo podemos asignarle dos posibles valores: F para femenino y M para masculino. Los símbolos F y M son etiquetas arbitrarias y no existe un orden entre ellas ni podemos realizar operaciones aritméticas. La religión o la nacionalidad son también ejemplos de variables nominales.

Escala ordinal

En esta escala de medición los valores de la variable tienen un orden pero no se pueden hacer operaciones aritméticas entre estos valores pues no hay una noción de distancia entre ellos. Por ejemplo, para calificar las características de un objeto podemos suponer los siguientes valores: 0=Pésimo, 1=Malo, 2=Regular, 3=Bueno, 4=Excelente. En este caso la escala de medición es ordinal pues existe un orden entre sus valores, pero no se puede decir, por ejemplo, que dos valores regulares, $2 + 2$, hacen un valor excelente, 4.

Escala de intervalo

En este tipo de escala existe un orden entre los valores de la variable y existe además una noción de distancia aunque no se pueden realizar operaciones ni existe el valor natural cero. Por ejemplo, suponga que los valores de una cierta variable están dados por los días del mes, entre el día 10 y el día 20 hay una distancia de diez días, pero no se puede decir que el día 20 es dos veces el día 10. La temperatura es otro ejemplo de este tipo de variable, el posible valor cero depende de la escala que se use para medir la temperatura (Celsius, Kelvin, Fahrenheit, etc.).

Escala de razón

En una escala de razón la magnitud tiene un sentido físico y existe el cero absoluto. Por ejemplo, la variable edad en años estudiada en una población humana o la corriente eléctrica que pasa por cable conductor.

Debemos mencionar, sin embargo, que la clasificación de variables en alguna de las escalas de medición mencionadas antes puede no ser un asunto sencillo, dicha clasificación dependerá del uso que de tal variable haga el observador o persona que lleva a cabo el estudio. A continuación se muestran algunos ejemplos de variables con distintas escalas de medición.

EJEMPLO 2.1. Para los siguientes ejemplos determinaremos si la variable indicada es cualitativa o cuantitativa. Si es cualitativa, diremos si su escala de medición es nominal u ordinal. Si es cuantitativa, especificaremos si es discreta o continua, y diremos su escala de medición.

a) Apreciación o punto de vista acerca de una idea o propuesta usando la escala: de acuerdo, en desacuerdo, indiferente.
   Esta es una variable cualitativa nominal.
b) Evaluación de un curso de acuerdo a las siguientes categorías: acreditado, no acreditado, ausente.
   Esta es una variable cualitativa ordinal.
c) Número de hijos en una familia.
   Esta es una variable cuantitativa, discreta, de razón.
d) Salario en unidades monetarias.
   Esta es una variable cuantitativa, discreta, de razón.
e) Nivel de salario en alguna de las categorías: bajo, medio, alto.
   Esta es una variable cualitativa ordinal.
f) Longitud de un tornillo.
   Esta es una variable cuantitativa, continua, de razón.
g) Preferencia sexual.
   Esta es una variable cualitativa nominal.
h) Lugar de nacimiento de una persona.
   Esta es una variable cualitativa nominal.
i) Peso de un bebé recién nacido.
   Esta es una variable cuantitativa, continua, de razón.

En el presente texto consideraremos principalmente variables cuantitativas.

3. Estadística descriptiva

Supongamos que tenemos un conjunto de datos numéricos $x_1, \ldots, x_n$, que representan mediciones de alguna variable cuantitativa de interés en un experimento aleatorio. Para

conocer algunas características globales de esta variable se pueden calcular ciertas medidas de tendencia central como la media, moda y mediana; y también otras medidas llamadas de dispersión como la varianza, la desviación estándar y el rango. Definiremos estos conceptos a continuación, así como algunos otros relativos a la representación gráfica de datos.

Media

La media o media aritmética de una colección de datos numéricos $x_1, \ldots, x_n$, denotada por $\bar{x}$, es simplemente el promedio aritmético:

$$\bar{x} = \frac{x_1 + \cdots + x_n}{n}.$$

Moda

La moda es el valor numérico que aparece con mayor frecuencia en el conjunto de datos. Si ningún valor se repite, se dice que no hay moda. Si existe un único valor con mayor número de repeticiones, entonces a ese valor se le llama la moda, y el conjunto de datos se dice que es unimodal. Pueden existir, si embargo, dos o mas valores que sean los que mayor número de veces se repiten, en tal caso la moda consta de esos valores y se dice que el conjunto de datos es bimodal si hay dos valores con mayor número de repeticiones. Si hay mas de dos valores con mayor número de repeticiones se dice que el conjunto de datos es multimodal.

Mediana

Para calcular la mediana de un conjunto de datos numéricos procedemos como sigue: se ordena la muestra de números $x_1, \ldots, x_n$ de menor a mayor incluyendo repeticiones y se obtiene la muestra ordenada $x_{(1)}, \ldots, x_{(n)}$, en donde $x_{(1)}$ denota el dato más pequeño y $x_{(n)}$ es el dato más grande. La mediana, denotada por $\tilde{x}$, se define como sigue:

$$\tilde{x} = \begin{cases} \frac{1}{2}\left(x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}\right) & \text{si } n \text{ es par,} \\ x_{(\frac{n+1}{2})} & \text{si } n \text{ es impar.} \end{cases}$$

Esta definición establece que cuando tenemos un número impar de datos, la mediana es el dato ordenado que se encuentra justo a la mitad. Y cuando tenemos un número par de datos, la mediana se calcula promediando los dos datos ordenados que están enmedio.

EJEMPLO 2.2. Calcularemos la media, moda y mediana de cada uno de los siguientes conjuntos de datos.

a) $-5, 2, -3, -2, 4, 4, 2, 5$.
La media es 0.875, las modas son 2 y 4, y la mediana es 2.

b) 3.5, 2, 7.5, 3.5, 1.5, 4.5, 7.5.
La media es 4.285, las modas son 3.5 y 7.5, y la mediana es 3.5.
c) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.
La media es 5.5, no hay moda, y la mediana es 5.5.

Para una variable aleatoria X, discreta o continua, con una cierta distribución de probabilidad, se le llama mediana al valor real m tal que

$$P(X \leq m) \geq 1/2 \quad \text{y} \quad P(X \geq m) \geq 1/2,$$

es decir, el valor m es tal que la probabilidad acumulada hasta m es por lo menos $1/2$ y al mismo tiempo la probabilidad después de m es también por lo menos $1/2$.

Definiremos ahora algunas medidas que ayudan a cuantificar la dispersión que puede presentar un conjunto de datos numéricos $x_1, \ldots, x_n$.

Varianza y desviación estándar

La varianza, denotada por s^2, se define como sigue

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

en donde $\bar{x}$ es la media muestral definida antes. La desviación estándar es la raíz cuadrada positiva de s^2 y se le denota naturalmente por s.

Rango

El rango es el dato más grande menos el dato más pequeño, es decir, es la diferencia $x_{(n)} - x_{(1)}$.

EJEMPLO 2.3. Calcularemos la varianza, la desviación estándar y el rango de los conjuntos de datos del ejemplo anterior.

a) $-5, 2, -3, -2, 4, 4, 2, 5$.
La varianza es 13.839, la desviación estándar es 3.720, y el rango es 10.
b) 3.5, 2, 7.5, 3.5, 1.5, 4.5, 7.5.
La varianza es 5.821, la desviación estándar es 2.412, y el rango es 6.
c) 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.
La varianza es 9.16, la desviación estándar es 3.027, y el rango es 9.

Gráfica de pastel (*Pie chart*)

Esta es una gráfica circular dividida en sectores que permite comparar visualmente las frecuencias relativas o porcentajes de los valores observados de una variable. El ángulo en grados de un sector con frecuencia o porcentaje p_i está dado por la fórmula $\alpha_i = 360 \times p_i$. Por ejemplo, un equipo de fútbol que en la temporada ha jugado 15 partidos, de los cuales ha perdido 5 (33.3 %), empatado 3 (20 %), y ganado 7 (46.66 %), puede representar su desempeño mediante la gráfica que se muestra en la Figura 2.2.

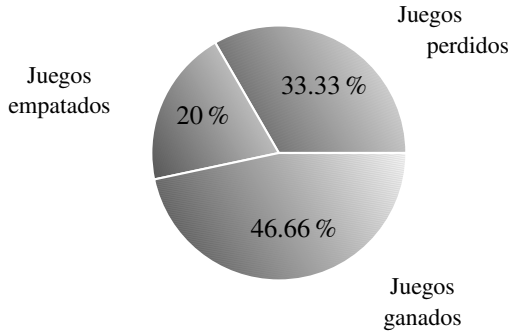


FIGURA 2.2. Gráfica de pastel.

Histogramas

Un histograma es una gráfica de barras en donde se dibuja una barra sobre cada uno de los valores observados de una variable y la altura de la barra indica la frecuencia con la que se observó el valor en cuestión.

EJEMPLO 2.4. Nos interesa estudiar el número de goles que un equipo particular de fútbol anota. Para ello se llevó el registro de los goles que el equipo anotó en sus últimos 32 partidos. Los resultados se resumen en la siguiente tabla:

Goles	0	1	2	3	4	5	6
Frecuencia	6	7	10	5	2	1	1

Una manera gráfica de representar estas frecuencias es a través del histograma que se muestra en la Figura 2.3. Visualmente puede ser más claro la comparación de las frecuencias registradas mediante diagramas de este estilo.

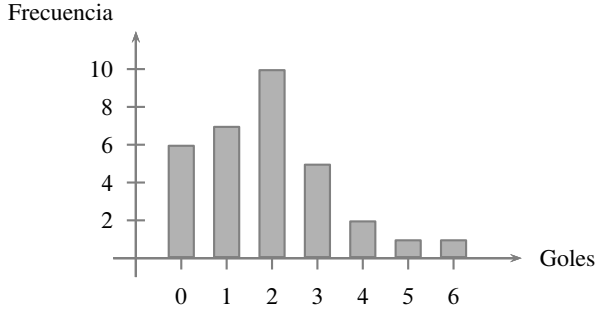


FIGURA 2.3

Así, se considera que se tienen 6 observaciones del valor 0, 7 observaciones del valor 1, 10 observaciones del valor 2, etcétera. Si f_i denota la frecuencia del valor i,

podemos entonces calcular la media de este conjunto de datos como sigue:

$$\begin{aligned}\bar{x} &= \frac{0\,f_0 + 1\,f_1 + \cdots + 6\,f_6}{\text{Total de observaciones}} \\ &= \frac{0(6) + 1(7) + \cdots + 1(6)}{32} \\ &= \frac{61}{32} \\ &= 1.90.\end{aligned}$$

Es decir, este equipo ha anotado un promedio de 1.90 goles por partido en sus últimos 32 juegos. La moda en este caso es el número de goles con mayor frecuencia y es 2. Ordenando todos los datos de menor a mayor incluyendo repeticiones, se encuentra que la mediana es

$$\tilde{x} = \frac{1}{2}(x_{(16)} + x_{(17)}) = 2.$$

La desviación estándar es, aproximadamente,

$$\begin{aligned}s^2 &= \frac{1}{31}\left((0-\bar{x})^2 f_0 + \cdots + (6-\bar{x})^2 f_6\right) \\ &= \frac{68.72}{31} \\ &= 2.21\,.\end{aligned}$$

Pueden calcularse también las frecuencias acumuladas como se muestra en la tabla que aparece abajo. La correspondiente gráfica de estas frecuencias acumuladas se muestra en la Figura 2.4.

Goles	0	1	2	3	4	5	6
Frec. acumulada	6	13	23	28	30	31	32

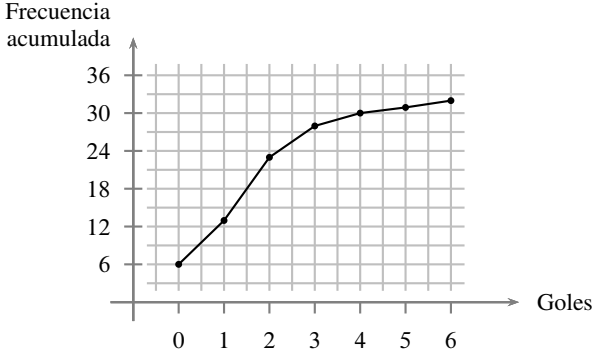


FIGURA 2.4

Cuando se dividen las frecuencias acumuladas entre el total de frecuencias, en este caso 32, se obtienen las probabilidades acumuladas. Esto se muestra en la tabla que aparece abajo. La gráfica correspondiente es un reescalamiento de la que se muestra en la Figura 2.4.

Goles	0	1	2	3	4	5	6
Prob. acumulada	6/32	13/32	23/32	28/32	30/32	31/32	32/32

Datos agrupados

En ocasiones los datos observados no se conocen con exactitud, solamente se sabe que un cierto número de observaciones se encuentran en un intervalo dado, por ejemplo $[a, b)$, y se tienen varios de estos intervalos con sus correspondientes frecuencias. En tal situación se dice que los datos están agrupados y las características numéricas que puedan calcularse serán necesariamente aproximadas. Por ejemplo, la información sobre la estatura de un grupo de 51 personas puede tenerse clasificada en la siguiente forma:

Estatura (cm)	150-159	160-169	170-179	180-189
Frecuencia	11	20	15	5

El objetivo de esta sección es mostrar la forma en que pueden calcularse algunas características numéricas de estos tipos de datos. Se puede obtener una colección de datos no agrupados a partir de una colección de datos agrupados calculando primero el punto medio de cada intervalo y considerar que dicho valor ha sido observado tantas veces como indica la frecuencia del intervalo. De esta forma pueden aplicarse los procedimientos para datos no agrupados y obtener sus propiedades y características numéricas de manera aproximada, aunque no todo es tan sencillo como parece, veamos algunos ejemplos:

a) Si se considera que la variable "Estatura" del ejemplo de la tabla anterior es continua, entonces el rango de valores $150 - 159$ corresponde al intervalo $[149.5, 159.5)$ y por lo tanto el punto medio de este intervalo es 155.
b) Considere la variable "Años cumplidos" para un grupo de personas. Si se considera que esta variable es continua, entonces el rango de valores $15 - 17$, por ejemplo, corresponde al intervalo $[15, 18)$ y el punto medio es 16.5 . Si se considera que la variable es discreta, entonces el rango $15 - 17$ corresponde al conjunto $\{15, 16, 17\}$ y el punto medio o promedio es 16.

Así, es conveniente especificar la clasificación de cada variable y establecer el conjunto de valores que corresponde a cada agrupación.

Histogramas. Puesto que los intervalos o agrupaciones no son necesariamente de la misma longitud o del mismo número de elementos, para cada grupo se deben calcular las frecuencias relativas a esta longitud, daremos esta definición en el caso continuo:

$$\text{frec. relativa} = \frac{\text{frecuencia del intervalo}}{\text{longitud del intervalo}}. \tag{9}$$

Se construye entonces un histograma en donde cada rectángulo tiene como base la longitud del intervalo, como altura su frecuencia relativa y centrado en el punto medio del intervalo. Observe que el intervalo con mayor frecuencia no es necesariamente el que tiene el rectángulo más alto pues la altura es relativa a la longitud del intervalo. Cuando todos los intervalos tienen la misma longitud, el intervalo de mayor frecuencia es efectivamente el más alto. Como un ejemplo consideremos la siguiente modificación de la colección de datos agrupados de la variable "Estatura", en donde la variable se considera continua y los intervalos tienen todos longitud distinta:

Estatura (cm)	150	151-152	153-155
Long. intervalo	1	2	3
Frecuencia	11	20	27
Frecuencia relativa	11	10	9

El lector puede observar que se ha diseñado este ejemplo para mostrar de manera dramática que las frecuencias relativas mantienen el orden contrario de las frecuencias absolutas. Relativo a la longitud del intervalo, hay mas observaciones en el primer intervalo. El histograma se muestra en la Figura 2.5.

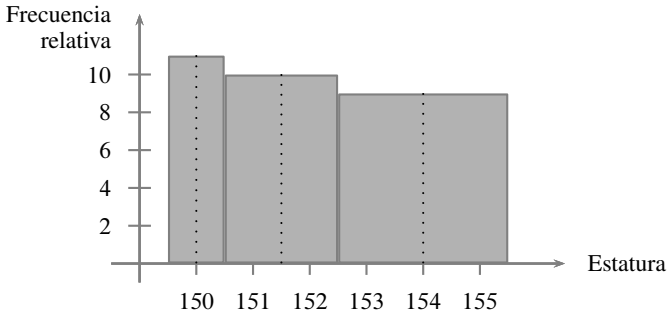


FIGURA 2.5

Media. Se calcula el valor medio x_i de cada intervalo o clase y se considera que tal valor se ha observado tantas veces como indica la frecuencia f_i del intervalo i . Se calcula entonces el valor promedio de todos estos valores: si el total de observaciones es n, entonces

$$\bar{x} = \frac{1}{n} \sum_i x_i f_i \, .$$

Por ejemplo, para los datos agrupados

Estatura (cm)	150-159	160-169	170-179	180-189
Valor medio x_i	155	165	175	185
Frecuencia f_i	11	20	15	5

la media es

$$\frac{1}{51}\sum_i x_i\, f_i = 167.74\,.$$

Moda. La moda de clase es el intervalo cuya frecuencia relativa es mayor. El punto medio de dicho intervalo puede tomarse como la moda, aunque no es el único método para este cálculo. Por ejemplo, para los datos agrupados cuyo histograma aparece en la Figura 2.5, la moda es el valor 150. Como en el caso de datos no agrupados, puede haber también varias modas para datos agrupados.

Mediana. Una forma sencilla de dar una aproximación a esta cantidad es utilizar los puntos medios x_i de cada intervalo, repetir ese valor tantas veces como indica la frecuencia f_i del intervalo y utilizar la definición de mediana para el caso de datos no agrupados.

Varianza y desviación estándar. La varianza se puede calcular como sigue

$$s^2 = \frac{1}{n-1}\sum_i (x_i - \bar{x})^2\, f_i\,,$$

en donde $\bar{x}$ es la media definida antes y n es el total de observaciones. La desviación estándar es la raíz cuadrada positiva de s^2.

4. Vectores aleatorios

Esta sección contiene una breve introducción al tema de variables aleatorias multidimensionales o también llamadas vectores aleatorios. Hasta ahora hemos considerado conjuntos de datos numéricos $x_1, x_2, \ldots, x_n$, pero podemos pensar que el valor x_1 es la observación o valor numérico tomado por una variable aleatoria X_1, x_2 es la observación de una variable aleatoria X_2 y así sucesivamente. Esta idea nos lleva a considerar conjuntos de variables aleatorias

$$X_1, X_2, \ldots, X_n,$$

en lugar del conjunto de números particulares $x_1, x_2, \ldots, x_n$. De esta forma X_1 representa el resultado de la primera observación, X_2 es el resultado de la segunda observación, etcétera, y por lo tanto desde esta perspectiva no estaremos tratando con valores numéricos particulares sino con variables aleatorias, las cuales constituirán el elemento básico para hacer inferencias. A estos conjuntos de variables aleatorias los llamaremos vectores aleatorios, estudiaremos algunas de sus propiedades y conceptos relacionados en las siguientes secciones y ejemplificaremos su utilidad en lo que resta del libro.

Para hacer la escritura corta se consideran únicamente vectores aleatorios de dimensión dos, aunque las definiciones y resultados que se mencionan pueden extenderse fácilmente, en la mayoría de los casos, para vectores de dimensión superior. Para el

material que se presenta a continuación sería provechoso contar con algunos conocimientos elementales del cálculo diferencial e integral en varias variables, o por lo menos mantener la calma cuando parezca que los símbolos matemáticos empiecen a perder sentido.

Vector aleatorio

Un vector aleatorio de dimensión dos es un vector de la forma (X, Y) en donde cada coordenada es una variable aleatoria. De manera análoga podemos tener vectores aleatorios multidimensionales $(X_1, \ldots, X_n)$. Nuevamente diremos que un vector aleatorio es discreto, o continuo, si las todas las variables aleatorias que lo conforman lo son. Por simplicidad, consideraremos únicamente vectores aleatorios cuyas coordenadas son variables aleatorias todas discretas, o continuas, pero no mezclas de ellas. Un vector aleatorio (X, Y) puede considerarse como una función de Ω en $\mathbb{R}^2$ como se muestra en la Figura 2.6.

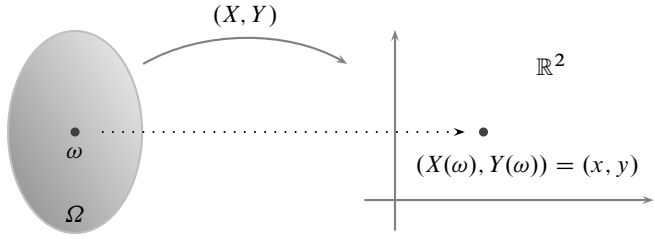


FIGURA 2.6. Vector aleatorio.

Es decir, el vector (X, Y) evaluado en ω es $(X, Y)(\omega) = (X(\omega), Y(\omega))$ con posible valor (x, y). Nuevamente observe que el vector con letras mayúsculas (X, Y) es el vector aleatorio, mientras que el vector con letras minúsculas (x, y) es un punto en el plano. Así, el vector $(X(\omega), Y(\omega))$ representa la información conjunta de dos respuestas o mediciones efectuadas a un mismo elemento ω del espacio muestral Ω. A veces la información de la que se dispone está agrupada de esta forma, en nuestro caso hemos mencionado únicamente dos variables aleatorias pero vectores de dimensión mayor son posibles.

EJEMPLO 2.5. Suponga que tenemos una población de mujeres y que la variable X toma el valor 1 cuando la mujer es fumadora y cero cuando no lo es. Sea Y la variable que registra el número de hijos de una mujer dada. Entonces el vector (X, Y) puede tomar los valores

$$(0, 0), (0, 1), (0, 2), \ldots$$
$$(1, 0), (1, 1), (1, 2), \ldots$$

Para una población particular de mujeres podríamos conformar una tabla con la frecuencia de cada una de estos valores del vector, por ejemplo,

x/y	0	1	2	3	4	5
0	2	5	8	3	2	0
1	10	9	2	0	1	1

EJEMPLO 2.6. Suponga que se cuenta con una población de personas que participan en un proceso de elección. El vector $(X(\omega), Y(\omega))$ puede representar el nivel económico y la preferencia electoral de un votante ω.

Estudiaremos a continuación algunas funciones asociadas a vectores aleatorios, las cuales son análogas al caso unidimensional estudiado antes.

Función de probabilidad conjunta

Vamos a estudiar primero el caso más sencillo que es el discreto. Consideremos un vector aleatorio discreto (X, Y) tal que la variable X toma valores en el conjunto $\{x_1, x_2, \ldots\}$, y la variable Y toma valores en $\{y_1, y_2, \ldots\}$. La función de densidad del vector (X, Y), que se denota por $f(x, y)$, se encuentra definida sobre $\mathbb{R}^2$ y toma valores en el intervalo $[0, 1]$ de acuerdo a la siguiente definición:

$$f(x, y) = \begin{cases} P(X = x, Y = y) & \text{si } (x, y) \in \{x_1, x_2, \ldots\} \times \{y_1, y_2, \ldots\}, \\ 0 & \text{otro caso.} \end{cases}$$

Es decir, el número $f(x, y)$ es la probabilidad de que la variable X tome el valor x y al mismo tiempo la variable Y tome el valor y. Tal función se llama también función de densidad conjunta de las variables X y Y, y para enfatizar este hecho a veces se escribe $f_{X,Y}(x, y)$, pero en general omitiremos los subíndices para hacer la notación más corta. Toda función $f(x, y)$ de la forma anterior cumple las siguientes dos propiedades, e inversamente, toda función definida sobre $\mathbb{R}^2$ que sea cero excepto en un conjunto discreto de parejas (x, y) que cumpla estas propiedades se llama función de probabilidad bivariada o conjunta, sin necesidad de contar con dos variables aleatorias previas que la definan.

a) $f(x, y) \geq 0$.
b) $\sum_{x,y} f(x, y) = 1$.

EJEMPLO 2.7. Considere el vector aleatorio discreto (X, Y) con función de densidad dada por la siguiente tabla:

x/y	0	1
-1	0.3	0.1
1	0.4	0.2

De este arreglo se entiende que la variable X toma valores en el conjunto $\{-1, 1\}$, mientras que Y toma valores en $\{0, 1\}$. Además las probabilidades conjuntas están dadas por las entradas de la tabla, por ejemplo, $P(X = -1, Y = 0) = 0.3$, esto es, la

probabilidad de que X tome el valor -1 y al mismo tiempo Y tome el valor 0 es 0.3. La misma información puede escribirse de la siguiente manera

$$f(x,y) = P(X = x, Y = y) = \begin{cases} 0.3 & \text{si } x = -1,\ y = 0, \\ 0.1 & \text{si } x = -1,\ y = 1, \\ 0.4 & \text{si } x = 1,\ y = 0, \\ 0.2 & \text{si } x = 1,\ y = 1, \\ 0 & \text{otro caso.} \end{cases}$$

Como todos estos valores son probabilidades, naturalmente son no negativos y todos ellos suman uno. Por lo tanto, $f(x,y)$ es efectivamente una función de probabilidad bivariada.

EJEMPLO 2.8. Encontraremos la constante c que hace a la siguiente una función de probabilidad conjunta.

$$f(x,y) = \begin{cases} cxy & \text{si } (x,y) \in \{1,2\} \times \{1,2\}, \\ 0 & \text{otro caso.} \end{cases}$$

Los posible valores del vector (X,Y) son $(1,1)$, $(1,2)$, $(2,1)$ y $(2,2)$, con probabilidades respectivas c, $2c$, $2c$ y $4c$. Como la suma de estas probabilidades debe ser uno, se llega a la ecuación $9c = 1$, de donde se obtiene que $c = 1/9$. Se deja como ejercicio al lector el graficar esta función.

Función de densidad conjunta

Veamos ahora la situación en el caso continuo. Sea (X,Y) un vector aleatorio continuo. Se dice que la función integrable y no negativa $f(x,y) : \mathbb{R}^2 \to [0,\infty)$ es la función de densidad del vector (X,Y) si para todo par (x,y) en $\mathbb{R}^2$ se cumple la igualdad

$$P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u,v)\, dv\, du.$$

Esta doble integral representa el volumen bajo la superficie dada por la función $f(u,v)$ sobre la región que se encuentra a la izquierda y abajo del punto (x,y). Toda función de densidad $f(x,y)$ de estas características satisface las siguientes dos propiedades que son análogas al caso discreto.

a) $f(x,y) \geq 0$.

b) $\displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\, dx\, dy = 1.$

Recíprocamente, decimos que una función $f(x,y) : \mathbb{R}^2 \to [0,\infty)$ es una función de densidad conjunta o bivariada si cumple con las dos condiciones arriba mencionadas.

EJEMPLO 2.9. La siguiente función es una de las funciones de densidad conjunta más sencillas. Sean $a < b$, $c < d$, y defina la función

$$f(x,y) = \begin{cases} \dfrac{1}{(b-a)(d-c)} & \text{si } a < x < b,\ c < y < d, \\ 0 & \text{otro caso,} \end{cases}$$

cuya gráfica aparece en la Figura 2.7. Se trata de una función constante en el rectángulo $(a,b) \times (c,d)$. Esta función es de densidad pues es no negativa e integra uno sobre $\mathbb{R}^2$. La doble integral sobre $\mathbb{R}^2$ es simplemente el volumen del paralelepípedo que se muestra en la Figura 2.7.

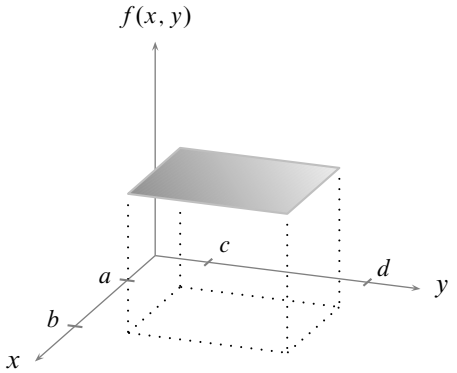


FIGURA 2.7. Función de densidad bivariada uniforme continua.

EJEMPLO 2.10. Comprobaremos que la siguiente función es de densidad.

$$f(x,y) = \begin{cases} x+y & \text{si } 0 < x, y < 1, \\ 0 & \text{otro caso.} \end{cases}$$

Claramente $f(x,y) \geq 0$ para cualquier $(x,y) \in \mathbb{R}^2$. Resta verificar que la función integra uno sobre el plano. El lector puede comprobar que

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = \int_0^1\int_0^1 (x+y)\,dxdy = \frac{1}{2} + \frac{1}{2} = 1.$$

EJEMPLO 2.11. Encontraremos la constante c para que la siguiente función sea de densidad.

$$f(x,y) = \begin{cases} cxy & \text{si } 0 < x < y < 1, \\ 0 & \text{otro caso.} \end{cases}$$

La constante c debe ser tal que la función $f(x,y)$ sea no negativa y que su integral sobre todo el plano sea uno. De esta última condición obtenemos que

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = \int_0^1\int_0^y cxy\,dxdy = \int_0^1 \frac{c}{2}\,y^3\,dy = \frac{c}{8}.$$

Por lo tanto $c = 8$.

Función de distribución conjunta

Además de la función de densidad, existe la función de distribución para un vector (X, Y), sea éste discreto o continuo, y su definición es muy semejante al caso unidimensional. La función de distribución del vector (X, Y), denotada por $F(x, y) : \mathbb{R}^2 \to [0, 1]$, se define para cualquier par de números reales (x, y) como sigue:

$$F(x, y) = P(X \leq x, Y \leq y).$$

La pequeña coma que aparece en el lado derecho de esta igualdad significa la intersección de los eventos $(X \leq x)$ y $(Y \leq y)$, es decir, el número $F(x, y)$ es la probabilidad del evento $(X \leq x) \cap (Y \leq y)$. Más precisamente, esta función debe escribirse como $F_{X,Y}(x, y)$, pero omitiremos los subíndices para mantener la notación simple. Siempre asociaremos la variable X con el valor x, y la variable Y con el valor y. A esta función se le conoce también con el nombre de función de acumulación de probabilidad del vector (X, Y), y también se dice que es la función de distribución conjunta de las variables X y Y. Enunciamos a continuación algunas propiedades que cumple toda función de distribución conjunta.

1. $\lim_{x,y \to \infty} F(x, y) = 1$.
2. $\lim_{x \to -\infty} F(x, y) = 0$. Mismo resultado cuando y tiende a menos infinito.
3. $F(x, y)$ es continua por la derecha en cada variable.
4. $F(x, y)$ es una función monótona no decreciente en cada variable.
5. Para cualesquiera números $a < b$, y $c < d$, se cumple la desigualdad

$$F(b, d) - F(a, d) - F(b, c) + F(a, c) \geq 0.$$

Observe que las primeras cuatro propiedades son análogas al caso unidimensional, y puede comprobarse geométricamente que la quinta propiedad corresponde a la probabilidad del evento $(a < X \leq b) \cap (c < Y \leq d)$. Recíprocamente, decimos que una función $F(x, y) : \mathbb{R}^2 \to [0, 1]$ es una función de distribución conjunta o bivariada si satisface las anteriores cinco propiedades. En el caso continuo supondremos que la función de distribución bivariada $F(x, y)$ puede expresarse de la siguiente forma:

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, dv\, du,$$

en donde $f(x, y)$ es una función no negativa y corresponde a la función de densidad bivariada asociada. El concepto de función de distribución bivariada puede extenderse al caso de vectores multidimensionales de la siguiente forma: la función de distribución del vector aleatorio $(X_1, \ldots, X_n)$ es la función $F(x_1, \ldots, x_n) : \mathbb{R}^n \to [0, 1]$ dada por

$$F(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n).$$

Las funciones $F(x, y)$ y $f(x, y)$ son equivalentes y en nuestro caso es siempre posible encontrar una a partir de la otra. Explicaremos este procedimiento a continuación.

De la función de densidad a la función de distribución. Conociendo la función de densidad $f(x, y)$, se puede encontrar la función de distribución $F(x, y)$ simplemente integrando en el caso continuo o sumando en el caso discreto. Para el caso continuo tenemos

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, dv\, du.$$

En el caso discreto se suman todos los valores de $f(u, v)$ para valores de u menores o iguales a x, y valores de v menores o iguales a y, es decir,

$$F(x, y) = \sum_{u \leq x} \sum_{v \leq y} f(u, v).$$

De la función de distribución a la función de densidad. Recíprocamente, puede encontrarse la función $f(x, y)$ a partir de $F(x, y)$ de la siguiente forma: en el caso continuo sabemos que $f(x, y)$ y $F(x, y)$ guardan la relación

$$F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, dv\, du,$$

y por el teorema fundamental del cálculo tenemos entonces que

$$f(x, y) = \frac{\partial^2}{\partial x\, \partial y} F(x, y).$$

En el caso discreto,

$$f(x, y) = F(x, y) - F(x-, y-).$$

Función de densidad marginal

Sea $f(x, y)$ la función de densidad del vector aleatorio continuo (X, Y). Se define la función de densidad marginal de la variable X como la siguiente integral

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy,$$

es decir, se integra simplemente respecto de la variable y para dejar como resultado una función que depende únicamente de x. Esta función resultante es la función de densidad marginal de X, y el subíndice 1 indica que se trata de la función de densidad marginal de la primera variable aleatoria del vector (X, Y). De manera completamente análoga, la función de densidad marginal de la variable Y se obtiene integrando ahora respecto de la variable x, es decir,

$$f_2(y) = \int_{-\infty}^{\infty} f(x, y)\, dx.$$

Nuevamente, el subíndice 2 hace referencia a que esta función es la función de densidad de la segunda variable aleatoria del vector (X, Y). En general, las funciones $f_1(x)$ y $f_2(y)$ son distintas, aunque hay ocasiones en que pueden ser iguales.

EJEMPLO 2.12. Sea (X, Y) un vector aleatorio continuo con función de densidad $f(x, y)$ dada por

$$f(x, y) = \begin{cases} 4xy & \text{si } x, y \in (0, 1), \\ 0 & \text{otro caso.} \end{cases}$$

Es sencillo verificar que esta función es efectivamente una función de densidad bivariada pues es no negativa e integra uno,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\, dxdy = \int_0^1 \int_0^1 4xy\, dxdy = 4 \cdot \frac{1}{2} \cdot \frac{1}{2} = 1.$$

Calcularemos ahora las funciones de densidad marginales $f_1(x)$ y $f_2(y)$. Para $x \in (0, 1)$,

$$f_1(x) = \int_{-\infty}^{\infty} f(x, y)\, dy = \int_0^1 4xy\, dy = 2x.$$

Por lo tanto,

$$f_1(x) = \begin{cases} 2x & \text{si } x \in (0, 1), \\ 0 & \text{otro caso.} \end{cases}$$

De manera análoga, o por simetría,

$$f_2(y) = \begin{cases} 2y & \text{si } y \in (0, 1), \\ 0 & \text{otro caso.} \end{cases}$$

Es inmediato comprobar que estas funciones son funciones de densidad univariadas. Observe que en este caso particular se cumple la identidad $f(x, y) = f_1(x) f_2(y)$, lo cual está relacionado con el importante concepto de independencia de v.a.s que mencionaremos más adelante.

La definición de función de densidad (o probabilidad) marginal para vectores discretos involucra una suma en lugar de la integral, por ejemplo,

$$f_1(x) = \sum_y f(x, y),$$

de manera análoga se define la función de densidad (o probabilidad) marginal $f_2(y)$. Es sencillo verificar que estas funciones de densidad marginales, tanto en el caso discreto como en el continuo, son efectivamente funciones de densidad (o probabilidad) univariadas.

EJEMPLO 2.13. Sea (X,Y) un vector aleatorio discreto con función de probabilidad $f(x,y)$ dada por

$$f(x, y) = \begin{cases} (x + 2y)/30 & \text{si } (x, y) \in \{1, 2, 3\} \times \{1, 2\}, \\ 0 & \text{otro caso.} \end{cases}$$

No es difícil comprobar que esta función es una función de probabilidad bivariada, es decir, es no negativa y suma uno,

$$\sum_{x=1}^{3}\sum_{y=1}^{2} \frac{x + 2y}{30} = 1.$$

Las funciones de probabilidad marginales $f_1(x)$ y $f_2(y)$ son

$$f_1(x) = \sum_{y=1}^{2} f(x,y) = \begin{cases} 8/30 & \text{si } x = 1, \\ 10/30 & \text{si } x = 2, \\ 12/30 & \text{si } x = 3, \\ 0 & \text{en otro caso.} \end{cases}$$

y

$$f_2(y) = \sum_{x=1}^{3} f(x,y) = \begin{cases} 12/30 & \text{si } y = 1, \\ 18/30 & \text{si } y = 2, \\ 0 & \text{en otro caso.} \end{cases}$$

Estas funciones son funciones de probabilidad univariadas.

Un poco más generalmente, la función de densidad marginal de la variable X_1 a partir de la función de densidad del vector $(X_1, \ldots, X_n)$ es, en el caso continuo,

$$f_1(x_1) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \ldots, x_n)\, dx_2 \cdots dx_n.$$

De manera análoga se obtiene la función marginal de cualquiera de las variables que componen el vector multidimensional. Y también de manera análoga se pueden calcular estas densidades marginales para vectores discretos de cualquier dimensión.

Función de distribución marginal

Sea (X, Y) un vector aleatorio, continuo o discreto, con función de distribución $F(x, y)$. La función de distribución marginal de la variable X se define como la función

$$F_1(x) = \lim_{y \to \infty} F(x, y),$$

y la correspondiente función de distribución marginal de la variable Y es la función

$$F_2(y) = \lim_{x \to \infty} F(x, y).$$

No es difícil comprobar que estas funciones de distribución marginales son efectivamente funciones de distribución univariadas. Todo lo mencionado en estas últimas secciones tiene como objetivo poder enunciar con precisión el concepto de independencia entre variables aleatorias. Veremos a continuación este importante concepto que será una hipótesis recurrente en los procedimientos de la estadística inferencial.

Independencia de dos variables aleatorias

Sean X y Y dos variables aleatorias discretas o continuas, con función de distribución conjunta $F(x, y)$ y con funciones de distribución marginales $F_1(x)$ y $F_2(y)$. Se dice que X y Y son independientes si se cumple la identidad

$$F(x, y) = F_1(x)\, F_2(y),$$

para cada uno de los valores de x y y. De la misma manera que en el caso de eventos independientes en donde tenemos que la ocurrencia de un evento no afecta la probabilidad de ocurrencia del otro, en el caso de independencia de variables aleatorias tenemos algo análogo: el evento de que una variable aleatoria tome un valor cualquiera no afecta a la probabilidad de que la otra variable aleatoria tome cualquiera de sus valores. Si $f(x, y)$, $f_1(x)$ y $f_2(y)$ denotan las correspondientes funciones de densidad o de probabilidad conjunta y marginales, la independencia entre X y Y también puede expresarse de la forma siguiente:

$$f(x, y) = f_1(x)\, f_2(y),$$

para cualesquiera valores de x y y.

EJEMPLO 2.14. Sean X y Y dos variables aleatorias discretas con función de probabilidad $f(x, y)$ dada por

$$f(x, y) = \begin{cases} 1/4 & \text{si } x, y \in \{0, 1\}, \\ 0 & \text{otro caso.} \end{cases}$$

Las funciones de probabilidad marginales son

$$f_1(x) = \begin{cases} 1/2 & \text{si } x \in \{0, 1\}, \\ 0 & \text{otro caso,} \end{cases} \qquad y \qquad f_2(y) = \begin{cases} 1/2 & \text{si } y \in \{0, 1\}, \\ 0 & \text{otro caso.} \end{cases}$$

Por lo tanto $f(x, y) = f_1(x) f_2(y)$ para cualesquiera números reales x y y, y se concluye entonces que X y Y son independientes.

A continuación enunciamos sin demostración un resultado que establece que la esperanza del producto de dos variables aleatorias independientes es el producto de las esperanzas.

PROPOSICIÓN 2.15. *Sean X y Y variables aleatorias independientes y con esperanza finita. Entonces*

$$E(XY) = E(X)\, E(Y).$$

En la siguiente sección extenderemos el concepto de independencia de dos variables aleatorias al caso de n variables aleatorias. Esta extensión será de particular utilidad en el tratamiento matemático que daremos de la estadística.

Independencia de n variables aleatorias

Sea $X_1, \ldots, X_n$ una colección de variables aleatorias con función de distribución conjunta $F(x_1, \ldots, x_n)$. Suponga que las respectivas funciones de distribución marginales son $F_1(x_1), \ldots, F_n(x_n)$. Se dice que estas variables aleatorias son independientes si para cualesquiera números reales $x_1, \ldots, x_n$ se cumple la igualdad

$$F(x_1, \ldots, x_n) = F_1(x_1) \cdots F_n(x_n).$$

Alternativamente, puede definirse la independencia en términos de la función de densidad como sigue

$$f(x_1, \ldots, x_n) = f_1(x_1) \cdots f_2(x_n).$$

Nuevamente esta igualdad debe verificarse para cualesquiera valores $x_1, \ldots, x_n$. Puede demostrarse que las dos condiciones anteriores son equivalentes. En el caso particular cuando las variables aleatorias son discretas, la condición de independencia se escribe de la forma siguiente: para cualesquiera números $x_1, \ldots, x_n$,

$$P(X_1 = x_1, \ldots, X_n = x_n) = P(X_1 = x_1) \cdots P(X_n = x_n).$$

EJEMPLO 2.16. Sean X y Y dos variables aleatorias continuas con función de densidad conjunta $f(x, y)$ dada por

$$f(x, y) = \begin{cases} e^{-x-y} & \text{si } x, y > 0, \\ 0 & \text{otro caso.} \end{cases}$$

Pueden calcularse las funciones de densidad marginales y encontrar que

$$f_1(x) = \begin{cases} e^{-x} & \text{si } x > 0, \\ 0 & \text{otro caso,} \end{cases} \qquad y \qquad f_2(y) = \begin{cases} e^{-y} & \text{si } y > 0, \\ 0 & \text{otro caso.} \end{cases}$$

Se verifica entonces que $f(x, y) = f_1(x) f_2(y)$ para cualesquiera números reales x y y, y ello demuestra la independencia de las variables X y Y.

Adicionalmente, se dice que un conjunto infinito de variables aleatorias son independientes si cualquier subconjunto finito de ellas lo es. Este es el sentido en el que la sucesión infinita de variables aleatorias debe entenderse, por ejemplo, en el enunciado del teorema central del límite o la ley de los grandes números.

Covarianza

Sean X y Y dos variables aleatorias con esperanza finita y con función de densidad o de probabilidad conjunta $f(x, y)$. La covarianza entre X y Y es un número que se denota por $\text{Cov}(X, Y)$ y se define como la esperanza de la variable aleatoria $(X - E(X))(Y - E(Y))$.

DEFINICIÓN 2.17. La covarianza entre las variables aleatorias X y Y es el número

$$\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))].$$

Como veremos más adelante, la covarianza está estrechamente relacionada con otro concepto que se define para dos variables aleatorias llamado coeficiente de correlación, y para el cual se cuenta con una interpretación bastante clara. Dejaremos entonces la interpretación de la covarianza en términos del coeficiente de correlación. Explicaremos ahora la forma de calcular la covarianza según la definición anterior. Cuando X y Y son variables aleatorias discretas la covarianza se calcula de la forma siguiente

$$\text{Cov}(X, Y) = \sum_{x,y} (x - E(X))(y - E(Y))\, f(x, y),$$

en donde, observe, la suma es doble, se suma sobre todos los posibles valores x y también sobre todos los posibles valores y. En el caso cuando las variables aleatorias son continuas se tiene que

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E(X))(y - E(Y))\, f(x, y)\, dxdy.$$

Mencionaremos a continuación algunas propiedades generales de la covarianza. Desarrollando el producto que aparece en la definición de covarianza y aplicando la linealidad de la esperanza se encuentra que la covarianza puede calcularse también como indica la siguiente fórmula:

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

Por otro lado, a partir de la definición misma de covarianza, o a partir de la fórmula recién enunciada, es inmediato observar que la covarianza es simétrica, es decir, $\text{Cov}(X, Y) = \text{Cov}(Y, X)$. Otra propiedad interesante y fácil de obtener se encuentra cuando se calcula la covarianza entre una variable aleatoria X y ella misma, en este caso la covarianza se reduce a la varianza de X. En el caso cuando las variables X y Y son independientes tenemos que $\text{Cov}(X, Y) = 0$. El recíproco es en general falso, es decir, el hecho de que la covarianza sea cero no implica necesariamente que las variables aleatorias en cuestión sean independientes. Por último, recordemos que hemos mencionado que la varianza de la suma de dos variables aleatorias no es, en general, la suma de las varianzas, sin embargo se cuenta con la siguiente fórmula general la cual puede ser encontrada a partir de la definición de varianza.

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2\ \text{Cov}(X, Y).$$

En las aplicaciones, a menudo no se conoce la distribución verdadera del vector aleatorio (X, Y) pero puede contarse con una colección de observaciones o mediciones $(x_1, y_1), \ldots, (x_n, y_n)$. Se puede entonces calcular la así llamada covarianza muestral como sigue y que representa una estimación de la verdadera covarianza entre las variables aleatorias:

$$\text{Cov}(\{x\}, \{y\}) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})\, (y_i - \bar{y}),$$

en donde $\bar{x} = (x_1 + \cdots + x_n)/n$ y $\bar{y}$ se define de manera análoga. En particular,

$$\text{Cov}(\{x\}, \{x\}) = \text{Var}(\{x\}) = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

Coeficiente de correlación

Habiendo definido la covarianza podemos ahora dar la definición del coeficiente de correlación entre dos variables aleatorias. Supondremos que tales variables aleatorias tienen esperanza y varianza finitas.

DEFINICIÓN 2.18. El coeficiente de correlación entre las variables aleatorias X y Y se define como el número

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\,\text{Var}(Y)}}.$$

Al número $\rho(X,Y)$ se le denota también por $\rho_{X,Y}$, en donde ρ es la letra griega ro. El lector puede observar inmediatamente que la diferencia entre la covarianza y el coeficiente de correlación radica únicamente en que este último se obtiene al dividir la covarianza por el producto de las desviaciones estándar de las variables aleatorias. Puede demostrarse que este cambio de escala tiene como consecuencia que el coeficiente de correlación tome como valor máximo 1, y como valor mínimo -1, es decir,

$$-1 \leq \rho(X,Y) \leq 1.$$

Explicaremos ahora la interpretación del coeficiente de correlación. Cuando X y Y son tales que $\rho(X,Y) = 1$, entonces existen constantes a y b, con a positiva tales que $Y = aX + b$, es decir, se puede establecer una dependencia lineal directa entre las dos variables aleatorias. En el otro caso extremo, cuando $\rho(X,Y) = -1$, entonces nuevamente existen constantes a y b, pero ahora con a negativa, tales que $Y = aX + b$. De nuevo, se trata de una relación lineal entre las dos variables aleatorias pero ahora tal relación es inversa en el sentido de que cuando una de las variables aleatorias crece la otra decrece. De esta forma el coeficiente de correlación es una medida del grado de dependencia lineal entre dos variables aleatorias. Existen varias formas en que dos variables aleatorias pueden depender una de otra, el coeficiente de correlación no mide todas estas dependencias, únicamente mide la dependencia de tipo lineal. Así, hemos mencionado que cuando el coeficiente de correlación es $+1$, ó -1, la dependencia lineal es exacta. Como en el caso de la covarianza, puede demostrarse que si dos variables aleatorias son independientes, entonces el coeficiente de correlación es cero, y nuevamente, el recíproco es en general falso, es decir, la condición de que el coeficiente de correlación sea cero no es suficiente para garantizar que las variables aleatorias sean independientes, excepto en el caso cuando las variables tienen distribución conjunta normal.

Si en lugar de la distribución del vector aleatorio (X,Y) se cuenta con una serie de observaciones $(x_1,y_1),\ldots,(x_n,y_n)$ de este vector, entonces puede calcularse el así llamado coeficiente de correlación muestral como sigue:

$$\rho(\{x\},\{y\}) = \frac{\text{Cov}(\{x\},\{y\})}{\sqrt{\text{Var}(\{x\})\ \text{Var}(\{y\})}}.$$

Esta cantidad representa una estimación del coeficiente de correlación entre las variables aleatorias.

5. Muestras aleatorias y estadísticas

En la teoría estadística que desarrollaremos en esta parte del texto no consideraremos muestras particulares de datos numéricos $x_1, \ldots, x_n$, sino conjuntos de variables aleatorias $X_1, \ldots, X_n$. Estas variables aleatorias pueden, en particular, tomar los valores numéricos mencionados. De esta manera y de forma general, estaremos considerando cualesquiera resultados que se puedan obtener al llevar a cabo un proceso de muestreo bajo algunas hipótesis adicionales que mencionaremos a continuación.

DEFINICIÓN 2.19. Una muestra aleatoria (escribimos simplemente m.a.) es una colección de variables aleatorias $X_1, \ldots, X_n$ que son independientes e idénticamente distribuidas.

De este modo, cuando se diga, por ejemplo, que una muestra aleatoria es tomada de una población normal con media μ y varianza σ^2, ello significa que las variables aleatorias que forman la m.a. son independientes entre sí, y todas ellas tienen la misma distribución normal con los mismos parámetros. Una muestra aleatoria constituye el elemento básico para llevar a cabo inferencias estadísticas.

DEFINICIÓN 2.20. Una estadística es una función cualquiera de una muestra aleatoria $X_1, \ldots, X_n$, que no depende de parámetros desconocidos. Por lo tanto, una estadística es una variable aleatoria.

Veremos a continuación algunos ejemplos de estadísticas que serán usados con frecuencia más adelante.

EJEMPLO 2.21. Considere una muestra aleatoria $X_1, \ldots, X_n$. La función $\bar{X}$ definida como aparece abajo es un ejemplo de una estadística, y se le conoce con el nombre de media muestral.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

EJEMPLO 2.22. La varianza muestral, denotada por S^2, es otro ejemplo de una estadística pues es una función de una muestra aleatoria definida de la forma siguiente:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

EJEMPLO 2.23. Sean $X_1, \ldots, X_n$ y $Y_1, \ldots, Y_n$ dos muestras aleatorias del mismo tamaño aunque no necesariamente con la misma distribución. La covarianza muestral se define como

$$S_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})\,(Y_i - \bar{Y}).$$

En particular,

$$S_{XX} = S_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

EJEMPLO 2.24. Sean $X_1, \ldots, X_n$ y $Y_1, \ldots, Y_n$ dos muestras aleatorias del mismo tamaño aunque no necesariamente con la misma distribución. El coeficiente de correlación muestral se define como

$$R(X, Y) = \frac{S_{XY}}{\sqrt{S_{XX}\, S_{YY}}}.$$

EJEMPLO 2.25. Sea $X_1, \ldots, X_n$ una muestra aleatoria. Las siguientes variables aleatorias son ejemplos de estadísticas:

a) $X_{(n)} = \max\{X_1, \ldots, X_n\}$.
b) $X_{(1)} = \min\{X_1, \ldots, X_n\}$.
c) $X = X_{(n)} - X_{(1)}$.

6. Estimación puntual

Sea X una variable aleatoria de interés en un experimento aleatorio y suponga que X tiene una distribución de probabilidad con función de densidad $f(x; \theta)$, en donde θ es el parámetro o el conjunto de parámetros de la distribución. En esta nueva notación se hace énfasis en que la distribución depende de un parámetro denotado genéricamente por la letra griega θ y el cual consideraremos desconocido. Por ejemplo, si la distribución es $\exp(\lambda)$, entonces θ representa el parámetro λ, si la distribución es $\mathrm{N}(\mu, \sigma^2)$, entonces θ representa el vector de parámetros (μ, σ^2). El problema de estimación puntual consiste en encontrar un número, con base en una serie de observaciones de la variable aleatoria, que sirva como estimación del parámetro desconocido θ. Ilustraremos el problema con un par de ejemplos.

EJEMPLO 2.26. Se desea conocer la calidad de un lote de mil artículos. Dada la imposibilidad de someter a prueba a todos ellos, se escogen 20 artículos al azar obteniéndose los siguientes resultados:

$$0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 0,$$

en donde cero indica que el artículo no pasó el control de calidad y uno indica que el artículo pasó el control de calidad. Suponga que X es la variable que indica si un artículo escogido al azar de la población completa pasa o no pasa el control de calidad. Entonces X tiene una distribución $\mathrm{Ber}(p)$, en donde el parámetro p es desconocido. ¿Cómo podría estimarse el verdadero valor de p con base en los datos de la muestra?

EJEMPLO 2.27. El tiempo en minutos que un conjunto de 10 empleados escogidos al azar invierte en trasladarse de la casa al lugar de trabajo es: 30, 70, 65, 10, 25, 15, 5, 50, 20, 15. Suponga que tal variable puede modelarse mediante la distribución $\exp(\lambda)$. ¿Cómo podría estimarse el valor de λ con base en las observaciones realizadas?

De esta forma, habiendo supuesto una distribución de probabilidad para una variable aleatoria de interés en donde la distribución depende de un parámetro no especificado en su valor, el problema consiste en encontrar un mecanismo para estimar

el parámetro desconocido tomando como elementos una serie de observaciones de la variable aleatoria.

DEFINICIÓN 2.28. Un estimador puntual para el parámetro θ es una función de una muestra aleatoria $X_1, \ldots, X_n$ que se usa para estimar θ.

A un estimador del parámetro θ se le denota regularmente por $\hat{\theta}$ (se lee "teta circunflejo"). Observe que un estimador puntual es una estadística y puede escribirse como $\hat{\theta} = \hat{\theta}(X_1, \ldots, X_n)$. Veremos a continuación dos métodos para encontrar estimadores puntuales.

Método de momentos

Este método fue introducido por el estadístico inglés Karl Pearson (né Carl Pearson) a principios del siglo XX. Sea nuevamente $f(x;\theta)$ la función de densidad de una variable aleatoria X que depende de un parámetro desconocido θ. Recordemos que el k-ésimo momento poblacional de X es el número

$$E(X^k), \qquad k = 1, 2, \ldots$$

cuando esta esperanza existe. Ahora, dada una muestra aleatoria $X_1, \ldots, X_n$ de esta distribución, se define el k-ésimo momento muestral como

$$\frac{1}{n}\sum_{i=1}^{n} X_i^k, \qquad k = 1, 2, \ldots$$

El método de momentos para estimar el parámetro θ es muy sencillo: consiste en igualar los momentos muestrales con los correspondientes momentos poblacionales, y resolver esta ecuación o sistema de ecuaciones para el parámetro θ cuando ello sea posible. Veamos algunos ejemplos.

EJEMPLO 2.29 (Estimación de un parámetro). Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de la distribución Ber(p). La estimación del parámetro p por el método de momentos consiste en igualar el primer momento poblacional (o primer momento de la distribución) p con el primer momento muestral dado por $\bar{X}$. Esta igualación produce directamente la respuesta:

$$\hat{p} = \bar{X}.$$

El parámetro desconocido p puede representar la probabilidad de una de las caras en una moneda, o la proporción de votantes a favor de una preferencia electoral, o la proporción de artículos defectuosos producidos por una determinada maquinaria, entre otras varias situaciones.

EJEMPLO 2.30 (Estimación de un parámetro). Considere una variable aleatoria continua X con función de densidad

$$f(x) = \begin{cases} \theta\, x^{\theta-1} & \text{si } 0 \leq x \leq 1, \\ 0 & \text{otro caso,} \end{cases}$$

en donde $\theta > 0$ es un parámetro. Usaremos el método de momentos para estimar este parámetro. Puede demostrar que $E(X) = \theta/(1+\theta)$. Igualando este promedio poblacional con el promedio muestral $\bar{X}$ se tiene la ecuación

$$\frac{\theta}{1+\theta} = \bar{X},$$

de donde se obtiene el estimador

$$\hat{\theta} = \frac{\bar{X}}{1-\bar{X}}, \qquad \bar{X} \neq 1.$$

EJEMPLO 2.31 (Estimación de dos parámetros). Nuevamente estimaremos los parámetros μ y σ^2 de una distribución normal. Esta vez usaremos el método de momentos. Como necesitamos estimar dos parámetros usamos los dos primeros momentos. El primer y segundo momento poblacionales son $E(X) = \mu$ y $E(X^2) = \sigma^2 + \mu^2$, respectivamente. El primer y segundo momento muestrales son $\frac{1}{n}\sum_{i=1}^{n} x_i$ y $\frac{1}{n}\sum_{i=1}^{n} x_i^2$, respectivamente. La igualación respectiva produce el sistema de ecuaciones

$$\begin{aligned} \mu &= \frac{1}{n}\sum_{i=1}^{n} x_i, \\ \sigma^2 + \mu^2 &= \frac{1}{n}\sum_{i=1}^{n} x_i^2. \end{aligned}$$

La primera ecuación es explícita mientras que la segunda ecuación se puede reescribir como sigue

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \mu^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (\frac{1}{n}\sum_{i=1}^{n} x_i)^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \mu)^2.$$

En este caso los estimadores por el método de momentos coinciden con los estimadores máximo verosímiles.

Método de máxima verosimilitud

Este método fue ampliamente popularizado a través de los trabajos que publicó el estadístico y genetista inglés Ronald Aylmer Fisher durante la segunda década del siglo XX, aunque la idea fundamental del método había sido usada con anterioridad por algunos matemáticos importantes como Gauss y Laplace. Explicaremos a continuación el método. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con función de densidad $f(x;\theta)$. Esto significa que todas las variables de la muestra aleatoria tienen función de densidad $f(x)$ que depende de un parámetro desconocido θ.

DEFINICIÓN 2.32. La función de verosimilitud de una muestra aleatoria $X_1, \ldots, X_n$, denotada por $L(\theta)$, se define como la función de densidad conjunta

$$L(\theta) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n; \theta).$$

Ronald A. Fisher
(Inglaterra 1890–1962)

La letra L proviene del término en inglés *likelihood*, que tradicionalmente se ha traducido como verosimilitud, aunque tal vez el término credibilidad sea más acertado. Aunque la función de verosimilitud depende de las observaciones $x_1, x_2, \ldots, x_n$, consideraremos a éstas como constantes y el que varía es el valor del parámetro desconocido θ. El método de máxima verosimilitud consiste en obtener el valor de θ que maximice la función de verosimilitud $L(\theta)$. La idea intuitiva es muy interesante: se debe encontrar el valor de θ de tal forma que los datos observados tengan máxima probabilidad de ocurrir. La probabilidad de ocurrencia está dada por la función de verosimilitud y por ello es que hay que maximizarla. El valor de θ en donde se alcanza el máximo se llama estimador de máxima verosimilitud o estimador máximo verosímil. Ilustraremos este método con algunos ejemplos.

EJEMPLO 2.33 (Estimación de un parámetro). Encontraremos el estimador máximo verosímil para el parámetro λ de una distribución exponencial. La función de verosimilitud es

$$L(\lambda) = f_{X_1}(x_1) \cdots f_{X_n}(x_n) = \lambda\, e^{-\lambda x_1} \cdots \lambda\, e^{-\lambda x_n} = \lambda^n\, e^{-\lambda n \bar{x}}.$$

Maximizar la función $L(\lambda)$ es equivalente a maximizar la función $\ln L(\lambda)$, pues la función logaritmo es continua y monótona creciente en su dominio de definición. Hacemos lo anterior porque la nueva función resulta más fácil de maximizar como veremos a continuación. Tenemos que

$$\ln L(\lambda) = n\, \ln \lambda - \lambda n \bar{x}.$$

Derivando respecto a λ e igualando a cero se llega a la ecuación

$$\frac{n}{\lambda} - n\bar{x} = 0,$$

de donde se obtiene $\lambda = 1/\bar{x}$. El estimador es entonces $\hat{\lambda} = 1/\bar{X}$.

Como se ha ilustrado en el ejemplo anterior, a menudo resulta más conveniente maximizar el logaritmo de la función de verosimilitud que la función de verosimilitud

misma. Idéntica situación se presentará en los ejemplos que a continuación se exponen. Por otro lado, es útil observar que el procedimiento de maximización utilizando derivadas es válido cuando el parámetro correspondiente puede tomar un continuo de valores, la función de verosimilitud sea diferenciable y cuando tal máximo exista.

EJEMPLO 2.34. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución geo(p). Comprobaremos que el estimador por máxima verosimilitud para el parámetro p está dado por $\hat{p} = 1/(1 + \bar{X})$. La función de verosimilitud es

$$\begin{aligned} L(p) &= f_{X_1}(x_1) \cdots f_{X_n}(x_n) \\ &= p\,(1-p)^{x_1} \cdots p\,(1-p)^{x_1} \\ &= p^n\,(1-p)^{n\bar{x}}. \end{aligned}$$

Tomando logaritmo,

$$\ln L(p) = n\,\ln p + n\bar{x}\,\ln(1-p).$$

Derivando respecto a p e igualando a cero se llega a la ecuación

$$\frac{n}{p} - \frac{n\bar{x}}{1-p} = 0,$$

de donde se obtiene $p = 1/(1 + \bar{x})$. El estimador es entonces $\hat{p} = 1/(1 + \bar{X})$.

EJEMPLO 2.35 (Estimación de dos parámetros). Encontraremos los estimadores de máxima verosimilitud para los parámetros μ y σ^2 de una distribución normal. Por definición, la función de verosimilitud es

$$\begin{aligned} L(\mu, \sigma^2) &= f_{X_1}(x_1) \cdots f_{X_n}(x_n) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_1-\mu)^2/2\sigma^2} \cdots \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x_n-\mu)^2/2\sigma^2} \\ &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i-\mu)^2}. \end{aligned}$$

Nuevamente el logaritmo de esta función es más sencillo de maximizar. Tenemos que

$$\ln L(\mu, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - \mu)^2.$$

Por lo tanto,

$$\begin{aligned} \frac{\partial}{\partial\mu} \ln L(\mu, \sigma^2) &= \frac{1}{\sigma^2}\sum_{i=1}^n (x_i - \mu), \\ \frac{\partial}{\partial\sigma^2} \ln L(\mu, \sigma^2) &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^n (x_i - \mu)^2. \end{aligned}$$

Igualando a cero ambas derivadas encontramos un sistema de dos ecuaciones con dos variables:

$$\begin{aligned} \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) &= 0, \\ -\frac{n}{2\sigma^2}+\frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 &= 0. \end{aligned}$$

De estas ecuaciones se obtiene $\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$, y $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2$. Por lo tanto los estimadores para los parámetros μ y σ^2 de una distribución normal por el método de máxima verosimilitud son

$$\begin{aligned} \hat{\mu} &= \frac{1}{n}\sum_{i=1}^{n} X_i, \\ \text{y} \quad \hat{\sigma}^2 &= \frac{1}{n}\sum_{i=1}^{n}(X_i-\hat{\mu})^2. \end{aligned}$$

EJEMPLO 2.36. Este ejemplo tiene el objetivo de mostrar que el procedimiento de maximizar la función de verosimilitud para encontrar el estimador máximo verosímil como lo hemos presentado en los ejemplos anteriores no siempre funciona. Considere una muestra aleatoria $X_1, \ldots, X_n$ de una distribución uniforme en el intervalo $[0, \theta]$, es decir, la función de densidad es $f(x) = 1/\theta$, para $0 \leq x \leq \theta$, en donde $\theta > 0$ es el parámetro a estimar. La función de verosimilitud es entonces $L(\theta) = 1/\theta^n$, para $x_1, \ldots, x_n \leq \theta$. Es fácil verificar que la función $L(\theta)$ no tiene un máximo en el intervalo $(0, \infty)$. ¿Qué hacer en este caso? Con base en el argumento intuitivo sugerido en la Figura 2.8, puede proponerse como estimador a la estadística $\hat{\theta} = \text{máx}\{X_1, \ldots, X_n\}$.

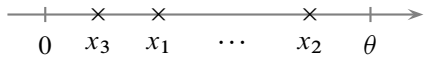


FIGURA 2.8

Insesgamiento

Puesto que un estimador $\hat{\theta}$ es una variable aleatoria que se utiliza para estimar el parámetro θ, es interesante saber si el valor promedio de $\hat{\theta}$ es θ. Esta sería una buena propiedad para un estimador y es lo que motiva la siguiente definición.

DEFINICIÓN 2.37. Un estimador $\hat{\theta}$ del parámetro θ es insesgado si

$$E(\hat{\theta}) = \theta.$$

Si $\hat{\theta}$ no es insesgado, entonces se dice que es sesgado y a la diferencia $E(\hat{\theta}) - \theta$ se le llama sesgo. De esta forma, un estimador puntual $\hat{\theta}$ es un estimador insesgado para el parámetro desconocido θ si, en promedio, el valor de $\hat{\theta}$ coincide con el valor desconocido de θ. En los siguientes ejemplos se muestra la forma en la que se verifica la propiedad de insesgamiento a pesar de desconocer el valor del parámetro θ.

EJEMPLO 2.38. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con media desconocida μ. Comprobaremos que la media muestral

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

es un estimador insesgado para el parámetro μ. Observe que $\bar{X}$ es el estimador $\hat{\theta}$ y μ es el parámetro desconocido θ. Por la propiedad de linealidad de la esperanza,

$$E(\hat{\mu}) = E(\bar{X}) = E(\frac{1}{n} \sum_{i=1}^{n} X_i) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} \sum_{i=1}^{n} \mu = \mu.$$

EJEMPLO 2.39. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con varianza desconocida σ^2. Recordemos que la varianza muestral es una estadística definida de la forma siguiente

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

En este caso el estimador es S^2 y el parámetro desconocido a estimar es σ^2. Esta estadística resulta ser un estimador insesgado para la varianza σ^2. Comprobar esta afirmación requiere de algunos cálculos que por cuestión de espacio omitimos, de modo que se deja al lector tal comprobación, lo único que es necesario hacer es desarrollar el cuadrado, usar la propiedad de linealidad de la esperanza y observar que

$$E(X_i X_j) = \begin{cases} \mu^2 & \text{si } i \neq j, \\ \sigma^2 + \mu^2 & \text{si } i = j. \end{cases}$$

7. Estimación por intervalos

En algunos casos es preferible no dar un número como estimación de un parámetro sino un intervalo de posibles valores. En esta sección se estudia brevemente el tema de estimación de parámetros usando intervalos. En este tipo de estimación se busca un intervalo de tal forma que se pueda decir, con cierto grado de confiabilidad, que dicho intervalo contiene el verdadero valor del parámetro desconocido. A este tipo de intervalos se les llama intervalos de confianza y fueron introducidos por el matemático y estadístico norteamericano de origen ruso-polaco Jerzy Neyman en 1937.

DEFINICIÓN 2.40. Sea $\alpha \in (0,1)$. Un intervalo de confianza para un parámetro desconocido θ de una distribución de probabilidad es un intervalo aleatorio de la forma $(\hat{\theta}_1, \hat{\theta}_2)$, en donde $\hat{\theta}_1$ y $\hat{\theta}_2$ son estadísticas (funciones de una muestra aleatoria) tales que

$$P(\hat{\theta}_1 < \theta < \hat{\theta}_2) = 1 - \alpha. \tag{10}$$

A las estadísticas $\hat{\theta}_1$ y $\hat{\theta}_2$ se les conoce como límites inferior y superior, respectivamente, del intervalo de confianza. Al número $1-\alpha$ se le conoce como grado o coeficiente de confianza. En general, se toma el valor de α cercano a 0 de tal forma que el grado de confianza, $1-\alpha$, es cercano a 1. En la práctica es común tomar $\alpha = 0.05$, de modo que el grado de confianza es $1-\alpha = 0.95$. Decimos entonces que el grado de confianza es del 95 %. Observe que las estadísticas $\hat{\theta}_1$ y $\hat{\theta}_2$ dependen de una muestra aleatoria $X_1, \ldots, X_n$, de modo que al tomar estas variables aleatorias distintos valores se generan distintos intervalos de confianza. Esta situación se ilustra en la Figura 2.9.

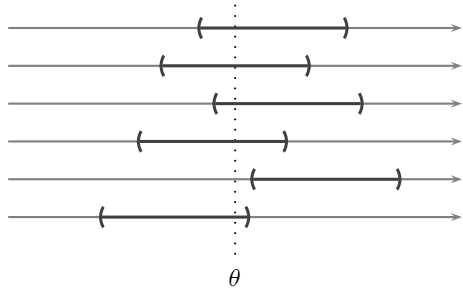


FIGURA 2.9. Intervalos.

Observe además que no es correcto decir "la probabilidad de que θ pertenezca al intervalo $(\hat{\theta}_1, \hat{\theta}_2)$ es $1-\alpha$", pues el parámetro θ no es el elemento aleatorio. En cambio, se dice "la probabilidad de que el intervalo $(\hat{\theta}_1, \hat{\theta}_2)$ contenga el valor de θ es $1-\alpha$". De esta forma se entiende que θ es constante, aunque desconocido, y el intervalo es el que cambia dependiendo de la muestra. Naturalmente el problema es encontrar $\hat{\theta}_1$ y $\hat{\theta}_2$ de tal forma que la igualdad (10) se cumpla. La idea general es encontrar una función de la muestra y del parámetro desconocido $q(X_1, \ldots, X_n; \theta)$, con distribución de probabilidad completamente conocida a partir de la cual pueda encontrarse un intervalo de confianza para θ. A tal función se le llama cantidad pivotal. En los siguientes párrafos mostraremos la forma de resolver este problema en algunos casos particulares.

Intervalo para la media de una distribución normal con varianza conocida

Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población normal con media desconocida μ y varianza conocida σ^2. Encontraremos un intervalo de confianza para el

parámetro μ. Como cada una de las variables de la muestra tiene distribución $\mathrm{N}(\mu, \sigma^2)$, la media muestral $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ tiene distribución $\mathrm{N}(\mu, \sigma^2/n)$. De modo que, estandarizando,

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathrm{N}(0, 1).$$

En esta situación, esta es la cantidad pivotal que nos ayudará a encontrar un intervalo de confianza para μ, y explicaremos a continuación el procedimiento. Para cualquier valor de $\alpha \in (0, 1)$ podemos encontrar un valor $z_{\alpha/2}$ en tablas de probabilidad normal estándar, véase la Figura 2.10, tal que

$$P(-z_{\alpha/2} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}) = 1 - \alpha.$$

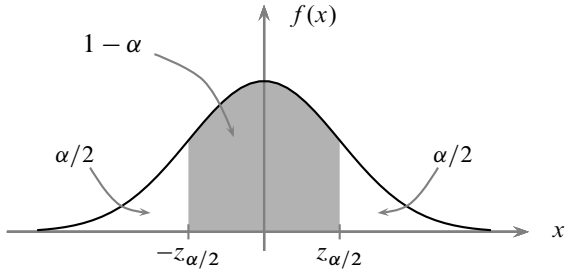


FIGURA 2.10

Despejando la constante desconocida μ se obtiene el siguiente resultado.

PROPOSICIÓN 2.41. *Un intervalo de confianza al* $(1-\alpha)100\,\%$ *para la media* μ *de una distribución normal con varianza conocida* σ^2 *está dado por la siguiente expresión*

$$P(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1 - \alpha. \tag{11}$$

De esta forma, el intervalo $(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}})$ es un intervalo de confianza para el parámetro desconocido μ, pues contiene a dicho parámetro con probabilidad $1 - \alpha$. Observe que todas las expresiones que aparecen en este intervalo son conocidas. Ilustraremos la aplicación de esta fórmula mediante un ejemplo.

EJEMPLO 2.42. Suponga que la vida promedio útil, medida en horas, de focos de 100 watts producidos por cierta compañía, puede ser modelada mediante una variable aleatoria con distribución normal de media μ y varianza σ^2. Suponga que la desviación estándar σ es conocida y es igual a 30 horas. El objetivo es encontrar un intervalo de confianza para la vida promedio útil μ de los focos producidos por esta compañía. Para ello se toma una muestra de 20 focos y mediante pruebas de laboratorio se determina la vida útil de cada uno de ellos. Los resultados $x_1, \ldots, x_{20}$ arrojan una media muestral

$\bar{x}$ de 1050 horas. Si consideramos un nivel de confianza del 95 %, es decir, $\alpha = 0.05$, de la tabla de probabilidad normal se encuentra que $z_{\alpha/2} = z_{0.025} = 1.96$, y entonces puede ahora calcularse el intervalo

$$\begin{aligned}(\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) &= (1050 - 1.96\,\frac{30}{\sqrt{20}},\ 1050 + 1.96 \times \frac{30}{\sqrt{20}}) \\ &= (1050 - 13.148,\ 1050 + 13.148) \\ &= (1036.852,\ 1063.148).\end{aligned}$$

De esta forma, con una confianza del 95 %, podemos afirmar que la vida promedio útil de este tipo de focos es de 1050 ± 13.148 horas.

Observe que la longitud del intervalo aleatorio que aparece en (11) es

$$2z_{\alpha/2}\frac{\sigma}{\sqrt{n}}.$$

De aquí pueden obtenerse varias observaciones interesantes:

a) La longitud del intervalo decrece conforme el tamaño de la muestra crece, es decir, mientras mayor información se tenga más preciso es el intervalo. En el límite cuando $n \to \infty$, el intervalo se colapsa en el estimador puntual $\bar{x}$.
b) Si la confianza requerida crece, es decir, si $1 - \alpha$ aumenta, entonces $z_{\alpha/2}$ crece, véase la Figura 2.10, y por lo tanto la longitud del intervalo también crece.
c) Si la dispersión de los datos es alta, es decir, desviación estándar σ grande, entonces la longitud del intervalo tiende a ser grande.

EJEMPLO 2.43. Un intervalo de confianza al 90 % para la media de una población normal con $\sigma = 5$ cuando se ha tomado una muestra de tamaño 25 cuya media muestral es 60 está dado por

$$\begin{aligned}(\,\bar{x} - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\,) &= (60 - 1.65\frac{5}{\sqrt{25}}, 60 + 1.65\frac{5}{\sqrt{25}}) \\ &= (58.35, 61.65).\end{aligned}$$

Intervalo para la media de una distribución normal con varianza desconocida

Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución normal con media desconocida μ pero ahora con varianza desconocida σ^2. El resultado teórico fundamental en la siguiente derivación es que la variable aleatoria

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

tiene una distribución t con $n - 1$ grados de libertad. Observe que esta es la distribución exacta de la variable T, sin importar el tamaño de la muestra y sobre todo, sin suponer que la varianza de la muestra es conocida. A partir de lo anterior podemos construir un intervalo de confianza para el parámetro desconocido μ de forma análoga al caso

normal mencionado antes. Para cualquier valor de $\alpha \in (0, 1)$ podemos encontrar un valor $t_{\alpha/2}$ en tablas de probabilidad de la distribución t de $n-1$ grados de libertad (véase la Figura 2.11) tal que

$$P(-t_{\alpha/2} < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2}) = 1 - \alpha.$$

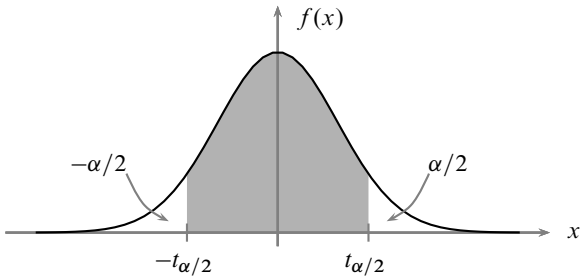


FIGURA 2.11

Despejando la constante desconocida μ de la ecuación anterior se obtiene el siguiente resultado.

PROPOSICIÓN 2.44. *Un intervalo de confianza para la media μ de una distribución normal está dado por la siguiente expresión*

$$P(\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}}) = 1 - \alpha. \tag{12}$$

De este modo, el intervalo $(\bar{X} - t_{\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2}\frac{S}{\sqrt{n}})$ es un intervalo de confianza para la media μ de una población normal sin suponer la varianza conocida. No lo hemos escrito de manera explícita en la fórmula anterior pero el valor $t_{\alpha/2}$ corresponde a la distribución t con $n-1$ grados de libertad. Para mayor precisión se escribe también $t_{\alpha/2,n-1}$.

Intervalo aproximado para la media de una distribución cualquiera

Sea $X_1, \ldots, X_n$ una muestra aleatoria de una distribución cualquiera con media desconocida μ. Supongamos que el tamaño n de la muestra es grande, por ejemplo, $n \geq 30$. Entonces, por el teorema central del límite, la variable aleatoria

$$Z = \frac{\bar{X} - \mu}{S/\sqrt{n}},$$

tiene una distribución aproximada normal estándar. Se puede encontrar un intervalo de confianza aproximado para el parámetro desconocido μ siguiendo el procedimiento

antes mencionado. Para cualquier valor de $\alpha \in (0,1)$ podemos encontrar un valor $z_{\alpha/2}$ en tablas de probabilidad normal estándar tal que

$$P(-z_{\alpha/2} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < z_{\alpha/2}) \approx 1-\alpha.$$

Despejando nuevamente la constante desconocida μ se obtiene

$$P(\bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}) \approx 1-\alpha.$$

De esta forma, el intervalo $(\bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}}, \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}})$ es un intervalo de confianza aproximado para el parámetro desconocido μ pues contiene a dicho parámetro con probabilidad $1-\alpha$. Observe nuevamente que todas las expresiones que aparecen en este intervalo son conocidas.

A manera de resumen se tiene la siguiente tabla.

Hipótesis	Intervalo para la media μ
Distribución normal varianza σ^2 conocida	$P(\bar{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}) = 1-\alpha.$
Distribución normal varianza σ^2 desconocida	$P(\bar{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}) = 1-\alpha.$
Cualquier distribución muestra grande, $n \geq 30$	Intervalo aproximado: $P(\bar{X} - z_{\alpha/2}\frac{S}{\sqrt{n}} < \mu < \bar{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}) \approx 1-\alpha.$

Intervalo aproximado para una probabilidad o proporción desconocida

Sea A un evento de interés de un experimento aleatorio cualquiera y sea p la probabilidad de dicho evento, la cual consideraremos desconocida. Deseamos encontrar un intervalo de confianza para estimar esta probabilidad. Supongamos que se realiza una sucesión de repeticiones independientes del experimento aleatorio y que nos interesa observar la ocurrencia (valor uno) o no ocurrencia (valor cero) del evento de interés en cada uno de estos ensayos. Sean $X_1, \ldots, X_n$ los resultados de estas observaciones, es decir, cada una de estas variables aleatorias tiene distribución Bernoulli de parámetro p. Sea $X = X_1 + \cdots + X_n$. Entonces X tiene distribución bin(n, p). Un estimador puntual para p es $\hat{p} = X/n$, en donde $E(\hat{p}) = p$ y $\text{Var}(\hat{p}) = p(1-p)/n$. Por el teorema central del límite, la variable aleatoria $\hat{p}$ tiene una distribución aproximada $\text{N}(p, p(1-p)/n)$. Por lo tanto puede encontrarse en tablas de probabilidad de la

distribución normal estándar un valor $z_{\alpha/2}$ tal que

$$P(-z_{\alpha/2} \leq \frac{\hat{p}-p}{\sqrt{p(1-p)/n}} \leq z_{\alpha/2}) \approx 1-\alpha. \tag{13}$$

La intención es despejar el valor de p en las desigualdades que definen a este evento, y presentamos a continuación tres formas en que tal tarea puede llevarse a cabo de manera aproximada.

a) Una simplificación al problema planteado consiste en substituir el denominador $p(1-p)/n$ por la estimación puntual $\hat{p}(1-\hat{p})/n$. Esto lleva a obtener el intervalo aproximado

$$P(\hat{p} - z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}) \approx 1-\alpha.$$

La longitud del intervalo es $\ell = 2\, z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$.

b) Otra alternativa para despejar el parámetro p en (13) es usar la desigualdad $p(1-p) \leq 1/4$, la cual es válida para valores de p entre cero y uno. De esta forma se tiene la estimación $\sqrt{p(1-p)/n} \leq 1/(2\sqrt{n})$, y en consecuencia despejar el parámetro p de (13) produce el siguiente intervalo aproximado con longitud mayor al presentado anteriormente

$$P(\hat{p} - z_{\alpha/2}/(2\sqrt{n}) \leq p \leq \hat{p} + z_{\alpha/2}/(2\sqrt{n})) \approx 1-\alpha.$$

La longitud del intervalo es $\ell = z_{\alpha/2}/\sqrt{n}$.

c) Como una tercera alternativa, observemos que el evento en cuestión puede escribirse como $(|\hat{p}-p| \leq z_{\alpha/2}\sqrt{p(1-p)/n})$. Elevando al cuadrado y desarrollando se llega a la desigualdad

$$p^2(1+z_{\alpha/2}^2/n) + p(-2\hat{p} - z_{\alpha/2}^2/n) + \hat{p}^2 \leq 0.$$

Las raíces de esta ecuación cuadrática en p son $p_1 = \hat{p}/(1+z_{\alpha/2}^2/n)$ y $p_2 = (\hat{p} + z_{\alpha/2}^2/n)/(1+z_{\alpha/2}^2/n)$. Por lo tanto, la ecuación cuadrática es no positiva cuando $p_1 \leq p \leq p_2$, es decir, se tiene entonces el intervalo de confianza no simétrico

$$P(\hat{p}/(1+z_{\alpha/2}^2/n) \leq p \leq (\hat{p} + z_{\alpha/2}^2/n)/(1+z_{\alpha/2}^2/n)) = 1-\alpha.$$

En este caso la longitud del intervalo es $\ell = z_{\alpha/2}^2/(n+z_{\alpha/2}^2)$.

EJEMPLO 2.45. Se han inspeccionado 92 artículos producidos por una cierta maquinaria de los cuales 7 resultaron con algún tipo de defecto. Encuentre un intervalo de confianza al 90 % para la proporción de artículos defectuosos producidos por la maquinaria en estudio.

Solución. Usaremos el intervalo aproximado dado por

$$P(\hat{p} - z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} \leq p \leq \hat{p} + z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}) \approx 1-\alpha,$$

en donde $n = 92$, $\hat{p} = 7/92 = 0.0760$, $\alpha = 0.1$ y por lo tanto $z_{\alpha/2} = 1.64$. Substituyendo estos valores en la fórmula mencionada se obtiene que

$$\begin{aligned} \hat{p} - z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} &= 0.0307 \\ \hat{p} + z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} &= 0.1214, \end{aligned}$$

así, con probabilidad 0.9, la verdadera proporción de artículos defectuosos producidos por la maquinaria está contenido en el intervalo $(0.0307, 0.1214)$, ello equivale a un porcentaje de artículos defectuosos entre el 3.07 % y el 12.14 %.

8. Pruebas de hipótesis

Empezaremos ilustrando la idea básica de una prueba de hipótesis mediante un ejemplo muy sencillo. Consideremos una situación en la que se efectúa sólo uno de los siguientes dos experimentos aleatorios: se lanza un dado equilibrado y se registra el número obtenido, o bien se lanza una moneda cinco veces y se registra el número de cruces totales que se obtienen (suponiendo que los lados de cada moneda se denominan cara y cruz). Supongamos que únicamente conocemos el resultado de uno u otro experimento y se nos pide determinar cuál de los dos experimentos se realizó con base en el número reportado.

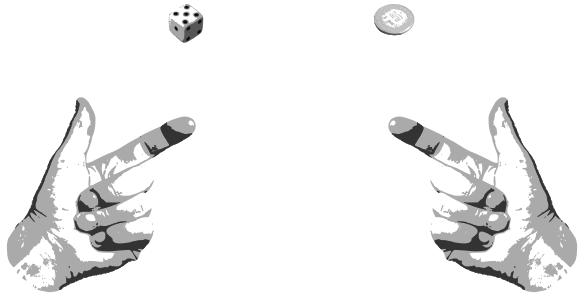


FIGURA 2.12

Tenemos entonces una situación de dos hipótesis:

$$H_0 : \text{“Dado”} \quad vs \quad H_1 : \text{“Moneda”},$$

en donde la primera hipótesis es H_0 y representa la situación en la que el número observado proviene del experimento de lanzar una dado, y la segunda hipótesis H_1 representa lo correspondiente al experimento de lanzar una moneda cinco veces. Como única información sobre este experimento tenemos entonces un número dentro del conjunto $\{0, 1, 2, 3, 4, 5, 6\}$, y con base en él debemos decidir si se llevó a cabo un experimento o el otro. Observe que si el número reportado es 0, entonces con seguridad

se realizó el experimento de la moneda. Si se reporta el número 6, entonces con seguridad el dado fue lanzado. ¿Qué decisión tomar para cualquier otro valor reportado? En la siguiente tabla se muestran las probabilidades de obtener los posibles números bajo cada uno de los dos experimentos.

x	0	1	2	3	4	5	6
H_0 : "Dado"	0	**1/6**	1/6	1/6	**1/6**	**1/6**	**1/6**
H_1 : "Moneda"	**1/32**	5/32	**10/32**	**10/32**	5/32	1/32	0

Debe ser claro que una estrategia natural es decidir por el experimento que tenga mayor probabilidad de producir el valor reportado. Estas probabilidades mayores se encuentran remarcadas en la tabla anterior. De esta forma se llega a la siguiente regla de decisión: si $x \in \mathcal{C} = \{0, 2, 3\}$, entonces se rechaza H_0 : "Dado" , en caso contrario se acepta H_0. Por razones naturales al conjunto $\mathcal{C}$ se le llama región de rechazo de la hipótesis H_0. La regla de decisión anterior es razonable, sin embargo no está libre de errores, por ejemplo, si $x = 2$, se decide por el experimento de la moneda, pero el resultado bien pudo provenir del dado. Igualmente, si $x = 1$, se decide por el dado pero es factible que el resultado haya sido obtenido por la moneda. Estas dos formas de cometer errores al efectuar una toma de decisión en una prueba de hipótesis se formalizan en las siguientes definiciones.

DEFINICIÓN 2.46 (Tipos de error).

a) El error tipo I se comete cuando se rechaza H_0 siendo ésta verdadera.
b) El error tipo II se comete cuando no se rechaza H_0 siendo ésta falsa.

Para el ejemplo anterior, las probabilidades de estos errores pueden ser calculadas del siguiente modo:

$$\begin{aligned} P(\text{"Error tipo I"}) &= P(\text{"Rechazar } H_0\text{"} \mid \text{"}H_0 \text{ es verdadera"}) \\ &= P(x \in \{0, 2, 3\} \mid \text{"Se usó el dado"}) \\ &= 0 + 1/6 + 1/6 \\ &= 1/3. \end{aligned}$$

Por otro lado,

$$\begin{aligned} P(\text{"Error tipo II"}) &= P(\text{"No rechazar } H_0\text{"} \mid \text{"}H_0 \text{ es falsa"}) \\ &= P(x \in \{1, 4, 5, 6\} \mid \text{"Se usó la moneda"}) \\ &= 5/32 + 5/32 + 1/32 + 0 \\ &= 11/32. \end{aligned}$$

Naturalmente, si se cambia la regla de decisión modificando la región de rechazo, entonces cambian las probabilidades de los errores. En un problema de decisión de este tipo se desea encontrar una regla de decisión que sea razonable y que tenga errores menores. Vamos ahora a estudiar pruebas de hipótesis en el contexto de la estimación de parámetros en las distribuciones de probabilidad.

DEFINICIÓN 2.47. Una hipótesis estadística o simplemente hipótesis es una afirmación o conjetura acerca de la distribución de una o mas variables aleatorias. Una hipótesis es simple si especifica por completo la distribución de probabilidad en cuestión, en caso contrario, la hipótesis se llama compuesta.

EJEMPLO 2.48.

a) Si X tiene una distribución bin(n, p), entonces la afirmación "$p = 0.2$" es una hipótesis.
b) Si X tiene una distribución N(μ, σ^2), entonces la afirmación "$\mu > 0$" es otro ejemplo de hipótesis estadística.

EJEMPLO 2.49.

a) Si X tiene una distribución exp(λ), entonces la afirmación "$\lambda = 5$" es una hipótesis simple.
b) Si X tiene una distribución N$(\mu, 1)$, entonces la afirmación "$\mu = 0$" es otro ejemplo de hipótesis simple.
c) Si X tiene una distribución Poisson(λ), entonces "$\lambda > 20$" es una hipótesis compuesta.
d) Si X tiene una distribución $\chi^2(n)$, entonces "$n \neq 5$" es otro ejemplo de una hipótesis compuesta.

En general, contrastaremos dos hipótesis de acuerdo al siguiente esquema y notación.

$$H_0 : \text{(hipótesis nula)} \quad vs \quad H_1 : \text{(hipótesis alternativa)}.$$

Tanto la hipótesis nula (H_0) como la hipótesis alternativa (H_1) pueden ser simple o compuesta. De este modo tenemos cuatro diferentes tipos de contraste de hipótesis: simple vs simple, simple vs compuesta, compuesta vs simple, y compuesta vs compuesta.

DEFINICIÓN 2.50. Una prueba de hipótesis es una regla para decidir si se acepta la hipótesis nula o se rechaza en favor de la hipótesis alternativa.

Como hemos mencionado, al tomar una decisión en una prueba de hipótesis se corre el riesgo de cometer errores. Rechazar la hipótesis nula cuando ésta es verdadera es el error tipo I y la probabilidad de cometer este tipo de error se le denota por la letra α. En cambio, no rechazar la hipótesis nula cuando ésta es falsa es el error tipo II y a la probabilidad de cometer este segundo tipo de error se le denota por la letra β. Estas definiciones de errores y sus probabilidades se resumen en la siguiente tabla.

	H_0 cierta	H_0 falsa
Rechazar H_0	Eror tipo I probabilidad α	Decisión correcta
No rechazar H_0	Decisión correcta	Error tipo II probabilidad β

La información para obtener una regla de decisión que nos lleve a rechazar o no rechazar un hipótesis estadística provendrá de una muestra aleatoria de la distribución de que se trate. Observe además que al aceptar una hipótesis no se afirma que ésta sea absolutamente cierta, sino simplemente que es consistente con los datos de la muestra aleatoria y la regla de decisión. Si la información de la muestra cambia, posiblemente la decisión de rechazar o no rechazar también cambie. Antes de presentar algunas pruebas de hipótesis particulares mencionaremos algunos términos adicionales que se usan en estos procedimientos.

DEFINICIÓN 2.51. Se le llama región crítica a la región de rechazo de H_0, y a la probabilidad de cometer el error tipo I, esto es α, se le llama tamaño de la región crítica. A esta probabilidad se le conoce también con el nombre de nivel de significancia.

FIGURA 2.13

Utilizaremos el siguiente ejemplo para ilustrar nuevamente las ideas generales y el procedimiento de llevar a cabo una prueba de hipótesis.

EJEMPLO 2.52. Suponga que tenemos una moneda y que deseamos saber si está equilibrada o no lo está. Para ello lanzamos la moneda 100 veces y con base en esta información habremos de tomar una decisión respecto a si la moneda tiene la misma probabilidad de caer cara o cruz. Por ejemplo, si de los cien lanzamientos se obtienen cincuenta cruces, entonces podemos decir que tenemos evidencias para creer que la moneda está equilibrada, aunque tal resultado puede también obtenerse con una moneda no equilibrada (error tipo II). Pero, ¿qué decisión tomar si únicamente se obtienen 45 cruces? ¿y si se obtienen 60 cruces? Es claro que en estos últimos casos la decisión no es tan inmediata.

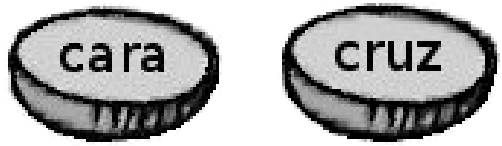


FIGURA 2.14

Denotemos por $X_1, X_2, \ldots, X_{100}$ los resultados de estos lanzamientos, en donde para el i-ésimo lanzamiento

$$X_i = \begin{cases} 1 & \text{si la moneda cae cruz,} \\ 0 & \text{si la moneda cae cara.} \end{cases}$$

Es decir, cada variable X_i tiene distribución Bernoulli de parámetro p, en donde p es la probabilidad desconocida de obtener cruz en cada lanzamiento. Deseamos llevar a cabo la prueba de hipótesis

$$H_0 : p = 1/2 \quad vs \quad H_1 : p \neq 1/2.$$

Denotemos por $\hat{p}$ a la media muestral $(X_1 + \cdots + X_{100})/100$. Por la ley de los grandes números, este estimador se acerca al verdadero valor de p cuando el número de lanzamientos es cada vez más grande, y por lo tanto es una aproximación de p. Cuando $\hat{p}$ diste mucho de $1/2$ es natural pensar que la moneda no está equilibrada. Es por ello que el criterio de decisión buscado consiste en rechazar la hipótesis H_0 cuando

$$|\, \hat{p} - 1/2 \,| \geq d,$$

para algún valor numérico de d que encontraremos a continuación estableciendo un valor particular para la probabilidad del error tipo I. En el caso cuando H_0 es cierta, pero se toma la decisión de rechazar H_0, se está en la situación de cometer el error tipo I, y establecemos que la probabilidad de que ello ocurra es α, igual 0.01 por ejemplo, es decir,

$$P(|\, \hat{p} - 1/2 \,| \geq d) = 0.01\,.$$

A partir de esta ecuación encontraremos el valor de d haciendo uso del hecho de que, cuando H_0 es verdadera, es decir, $p = 1/2$, el estimador $\hat{p}$ tiene distribución aproximada normal de media $1/2$, varianza $p(1-p)/n = (1/2)(1-1/2)/100 = 0.025$, y desviación estándar $\sqrt{0.025} = 0.05$. Por lo tanto, $(\hat{p} - 1/2)/0.05$ tiene distribución aproximada normal estándar. Con esta información podemos ahora encontrar el valor

de d.

$$\begin{aligned} 0.01 &= P(|\,\hat{p} - 1/2\,| \geq d) \\ &= 1 - P(|\,\hat{p} - 1/2\,| < d) \\ &= 1 - P(-d < \hat{p} - 1/2 < d) \\ &= 1 - P(\frac{-d}{0.05} < \frac{\hat{p} - 1/2}{0.05} < \frac{d}{0.05}) \\ &= 2(1 - \Phi(\frac{d}{0.05})). \end{aligned}$$

Es decir, $\Phi(d/0.05) = 0.995$. De la tabla de probabilidades de la distribución normal se encuentra que $d/0.05 = 2.57$ y por lo tanto $d = 0.128$. De esta forma la estadística de la prueba es la variable aleatoria $\hat{p}$, y cuando el valor de esta variable cumpla la condición $|\,\hat{p} - 1/2\,| \geq 0.128$, se rechaza la hipótesis H_0. Esta región de rechazo se puede escribir como la unión de los intervalos

$$[0, 1/2 - 0.128] \cup [1/2 + 0.128, 1],$$

o bien $[0, 0.372] \cup [0.628, 1]$, y se muestra gráficamente en la Figura 2.15.

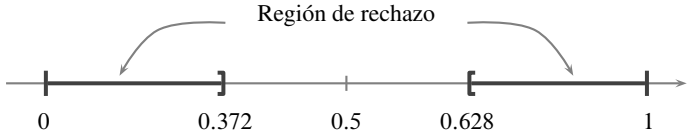


FIGURA 2.15

Así, si $\hat{p}$ es menor o igual a 0.372 o mayor o igual 0.628, decidimos que la diferencia entre $\hat{p}$ y $1/2$ no es debido a fluctuaciones azarosas, sino que es debido a que la moneda no está equilibrada y por lo tanto rechazamos H_0. La probabilidad de un error al tomar tal decisión es 0.01, de modo que se está tomando un riesgo del 1 % de clasificar una moneda equilibrada como no equilibrada. Por otro lado, podemos calcular la probabilidad del error tipo II de la siguiente forma: sea p_0 una probabilidad distinta de $1/2$, calcularemos la probabilidad del error tipo II dado que el verdadero valor de p es p_0, y escribimos esto como $\beta(p_0)$. Tenemos entonces que, suponiendo

$p = p_0$,

$$\begin{aligned}
\beta(p_0) &= P(\text{“No rechazar } H_0 \text{ cuando } p = p_0\text{”}) \\
&= P(\,|\hat{p} - 1/2| < d \mid p = p_0\,) \\
&= P(\,-d < \hat{p} - 1/2 < d \mid p = p_0\,) \\
&= P(\,1/2 - d < \hat{p} < 1/2 + d \mid p = p_0\,) \\
&\approx P(\,\frac{1/2 - d - p_0}{\sqrt{p_0(1-p_0)/100}} < Z < \frac{1/2 + d - p_0}{\sqrt{p_0(1-p_0)/100}}\,) \\
&= \Phi(\frac{1/2 + d - p_0}{\sqrt{p_0(1-p_0)/100}}) - \Phi(\frac{1/2 - d - p_0}{\sqrt{p_0(1-p_0)/100}}).
\end{aligned}$$

Observe que hemos aplicado el teorema central del límite al hacer la aproximación. De esta forma el error tipo II queda expresado como una función del valor de p_0 distinto de $1/2$.

Habiendo establecido las ideas principales y la notación que usaremos, podemos ahora mostrar la forma en la que se pueden encontrar algunas reglas de decisión para ciertas pruebas de hipótesis en estadística.

Prueba acerca de la media de una distribución normal con varianza conocida

Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población normal con media desconocida μ y varianza conocida σ^2. Sabemos que $\bar{X}$ tiene distribución $\mathrm{N}(\mu, \sigma^2/n)$. Por lo tanto,

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Sea μ_0 un número real particular. Deseamos contrastar las hipótesis

$$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0.$$

El problema es encontrar una regla para decidir cuándo rechazar H_0 en favor de H_1 con base en los datos de la muestra aleatoria. Cuando H_0 es cierta, esto es, cuando μ es efectivamente μ_0, tenemos que $\bar{X} \sim N(\mu_0, \sigma^2/n)$ y por lo tanto

$$\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1).$$

La estadística $Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$ es una medida natural de la distancia entre $\bar{X}$ (un estimador de μ), y su valor esperado μ_0 cuando H_0 es cierta. Es entonces razonable rechazar H_0 cuando la variable Z sea grande. Es por ello que tomamos como criterio de decisión rechazar H_0 cuando $|Z| \geq k$, para cierta constante k. ¿Cómo encontramos el número k? En una tabla de la distribución normal podemos encontrar un valor $z_{\alpha/2}$ tal que $P(|Z| \geq z_{\alpha/2}) = \alpha$, en donde α lo determina la persona que lleva a cabo la prueba de hipótesis, típicamente $\alpha = 0.1$. Véase la Figura 2.16. Este valor $z_{\alpha/2}$ es precisamente la constante k buscada pues con ello se logra que la región de rechazo sea de tamaño α.

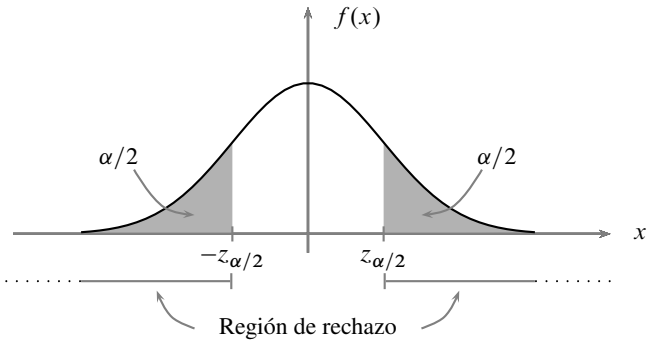


FIGURA 2.16

A la variable aleatoria Z se le llama la estadística de la prueba, y la prueba se denomina prueba de dos colas pues la región de rechazo consta de las dos colas de la distribución normal que se muestran en la Figura 2.16. Llevar a cabo esta prueba de hipótesis consiste simplemente en usar los datos de la muestra para encontrar el valor de Z, si resulta que $|Z| \geq z_{\alpha/2}$, entonces se rechaza H_0, en caso contrario no se rechaza H_0. En la siguiente tabla se muestra resumida la información de esta prueba.

Prueba:	$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0$
Estadística de prueba:	$Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$
Región de rechazo:	$\|Z\| \geq z_{\alpha/2}$, (prueba de dos colas)
Error tipo I:	α
Error tipo II:	$\Phi(z_{\alpha/2} + \frac{\mu_0-\mu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2} + \frac{\mu_0-\mu_1}{\sigma/\sqrt{n}})$, $\mu_1 \neq \mu_0$.

Vamos a comprobar la fórmula que aparece en la tabla anterior acerca del error tipo II. Sea μ_1 cualquier número tal que $\mu_1 \neq \mu_0$. Calcularemos la probabilidad del error tipo II dado que el verdadero valor de μ es μ_1.

$$\begin{aligned}
\beta(\mu_1) &= P(\text{“No rechazar } H_0 \text{ cuando } \mu = \mu_1\text{”}) \\
&= P(|Z| < z_{\alpha/2} \mid \mu = \mu_1) \\
&= P(|\frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}| < z_{\alpha/2} \mid \mu = \mu_1) \\
&= P(\mu_0 - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \bar{X} < \mu_0 + z_{\alpha/2}\frac{\sigma}{\sqrt{n}} \mid \mu = \mu_1) \\
&= P(-z_{\alpha/2} + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} < \frac{\bar{X} - \mu_1}{\sigma/\sqrt{n}} < z_{\alpha/2} + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}) \\
&= \Phi(z_{\alpha/2} + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}) - \Phi(-z_{\alpha/2} + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}).
\end{aligned}$$

EJEMPLO 2.53. En ciertas zonas de la ciudad y durante varios años se ha calculado el pago por el consumo de agua suponiendo un consumo promedio de 20,000 litros mensuales en cada casa. Para determinar si tal cantidad ha cambiado, se han medido los consumos mensuales de 15 casas obteniéndose los siguientes resultados: 23456, 18325, 21982, 22371, 13292, 25073, 22601, 20930, 18788, 19162, 21442, 23935, 20320, 19095, 17421. ¿Debe cambiar el consumo promedio mensual estimado para el cálculo de los pagos o permanecer igual? Suponga $\sigma = 2000$.

Solución. Supondremos un modelo normal para el consumo de agua con media desconocida μ y varianza conocida $\sigma^2 = (2000)^2$. Llevaremos a cabo la prueba de hipótesis

$$H_0 : \mu = 20,000 \quad vs \quad H_1 : \mu \neq 20,000.$$

Los datos proporcionados corresponden a los valores de una muestra aleatoria de tamaño 15, y haciendo el promedio de estos valores se obtiene una media muestral $\bar{X} = 20546.2$. La estadística de prueba toma entonces el valor

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{20546.2 - 20000}{2000/\sqrt{15}} = 1.0577.$$

Por otro lado, tomando $\alpha = 0.1$, de la tabla de probabilidades de la distribución normal se encuentra que $z_{\alpha/2} = 1.65$. Como no se cumple la condición $|Z| \geq z_{\alpha/2}$, la estadística de prueba Z cae fuera de la región de rechazo y por lo tanto no se rechaza la hipótesis H_0, es decir, no existen evidencias para afirmar que el consumo de agua por casa en la zona de estudio haya cambiado.

Puede también considerarse la prueba $H_0 : \mu = \mu_0$ contra $H_1 : \mu < \mu_0$ llamada prueba de cola inferior pues la región de rechazo consta de la cola izquierda de la distribución normal como se muestra en la Figura 2.17. Se rechaza la hipótesis H_0 sólo cuando los datos de la muestra son tales que $\bar{X}$ se encuentra muy a la izquierda de μ_0.

Prueba:	$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu < \mu_0$
Estadística de prueba:	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$
Región de rechazo:	$Z \leq -z_\alpha$, (prueba de cola inferior)
Error tipo I:	α
Error tipo II:	$1 - \Phi\left(-z_\alpha + \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}\right)$, $\mu_1 < \mu_0$

Las características de la prueba $H_0 : \mu = \mu_0$ contra $H_1 : \mu > \mu_0$ llamada prueba de cola superior se muestran en la siguiente tabla, y la región de rechazo se presenta en la Figura 2.18. Se rechaza la hipótesis H_0 sólo cuando la muestra provee evidencia de que $\bar{X}$ se encuentra muy a la derecha de μ_0.

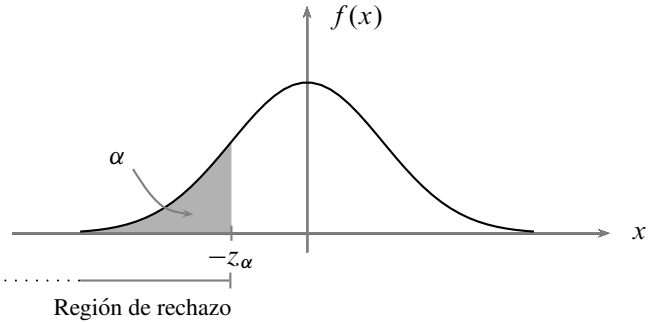


FIGURA 2.17

Prueba:	$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu > \mu_0$
Estadística de prueba:	$Z = \frac{\bar{X}-\mu_0}{\sigma/\sqrt{n}}$
Región de rechazo:	$Z \geq z_\alpha$, (prueba de cola superior)
Error tipo I:	α
Error tipo II:	$\Phi\left(z_\alpha + \frac{\mu_0-\mu_1}{\sigma/\sqrt{n}}\right)$, $\mu_1 > \mu_0$.

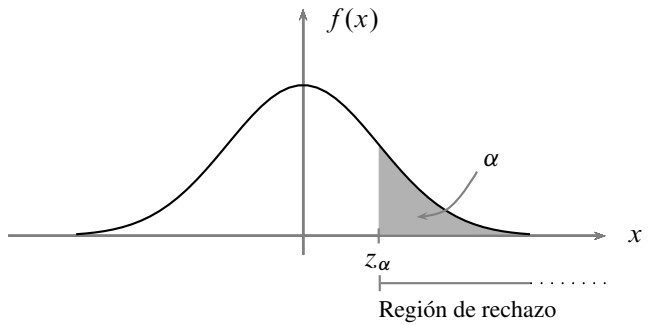


FIGURA 2.18

Prueba acerca de la media de una distribución normal con varianza desconocida

Sea nuevamente $X_1, \ldots, X_n$ una muestra aleatoria de una distribución normal con media desconocida μ, pero ahora con varianza desconocida σ^2. Dado un valor numérico para la constante μ_0, nos interesa nuevamente encontrar una regla de decisión para

llevar a cabo los siguientes tipos de pruebas de hipótesis

$$\begin{aligned} H_0 : \mu = \mu_0 \quad & vs \quad H_1 : \mu \neq \mu_0, \\ \cdots \quad & vs \quad H_1 : \mu < \mu_0, \\ \cdots \quad & vs \quad H_1 : \mu > \mu_0. \end{aligned}$$

El resultado teórico que hemos usado antes es que la variable aleatoria

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}},$$

tiene una distribución t con $n-1$ grados de libertad, en donde, recordemos, S^2 es la varianza muestral

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

El razonamiento para resolver este problema es completamente análogo al realizado en la sección anterior, sólo que ahora en lugar de usar la distribución normal se usa la distribución t. Por ejemplo, la prueba $H_0 : \mu = \mu_0$ vs $H_1 : \mu \neq \mu_0$, se denomina prueba de dos colas y nuevamente es razonable rechazar la hipótesis nula H_0 cuando la diferencia entre μ y μ_0 es grande. Los detalles de la prueba se encuentran en el recuadro siguiente y la región de rechazo se muestra gráficamente en la Figura 2.19.

Prueba:	$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu \neq \mu_0$
Estadística de prueba:	$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}$
Región de rechazo:	$\lvert t \rvert \geq t_{\alpha/2,n-1}$, (prueba de dos colas)
Error tipo I:	α
Error tipo II:	$F(t_{\alpha/2,n-1} + \frac{\mu_0 - \mu_1}{S/\sqrt{n}}) - F(-t_{\alpha/2,n-1} + \frac{\mu_0 - \mu_1}{S/\sqrt{n}})$, para $\mu_1 \neq \mu_0$.

El número $t_{\alpha/2,n-1}$ corresponde a aquel valor real tal que el área bajo la función de densidad de la distribución t con $n-1$ grados de libertad a la derecha de ese valor es $\alpha/2$, véase la Figura 2.19.

El error tipo I lo establece la persona que lleva a cabo la prueba y el error tipo II puede calcularse de manera análoga al caso cuando la varianza es conocida. Comprobaremos ahora la fórmula del error tipo II para la prueba de dos colas. Sea μ_1 cualquier número real distinto de μ_0. Calcularemos la probabilidad del error tipo II dado que el

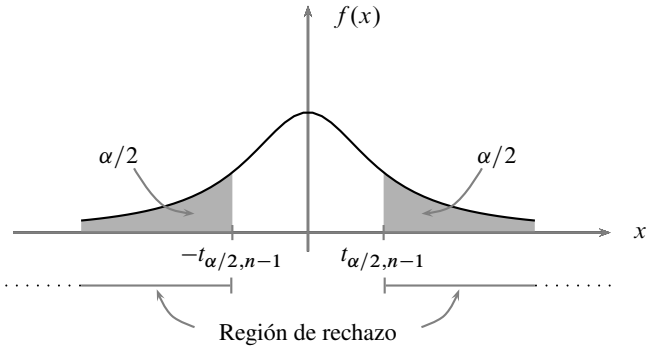


FIGURA 2.19

verdadero valor de la media es μ_1.

$$\begin{aligned}
\beta(\mu_1) &= P(\text{“No rechazar } H_0 \text{ cuando } \mu = \mu_1\text{”}) \\
&= P(|t| < t_{\alpha/2,n-1} \mid \mu = \mu_1) \\
&= P(|\frac{\bar{X}-\mu_0}{S/\sqrt{n}}| < t_{\alpha/2,n-1} \mid \mu = \mu_1) \\
&= P(\mu_0 - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \bar{X} < \mu_0 + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} \mid \mu = \mu_1) \\
&= P(-t_{\alpha/2,n-1} + \frac{\mu_0-\mu_1}{S/\sqrt{n}} < \frac{\bar{X}-\mu_1}{S/\sqrt{n}} < t_{\alpha/2,n-1} + \frac{\mu_0-\mu_1}{S/\sqrt{n}}) \\
&= F(t_{\alpha/2,n-1} + \frac{\mu_0-\mu_1}{S/\sqrt{n}}) - F(-t_{\alpha/2,n-1} + \frac{\mu_0-\mu_1}{S/\sqrt{n}}),
\end{aligned}$$

en donde F es la función de distribución t con $n-1$ grados de libertad. La prueba $H_0 : \mu = \mu_0$ *vs* $H_1 : \mu < \mu_0$ se llama nuevamente prueba de cola inferior y tiene las siguientes características:

Prueba:	$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu < \mu_0$
Estadística de prueba:	$t = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$
Región de rechazo:	$t \leq -t_{\alpha,n-1}$, (prueba de cola inferior)
Error tipo I:	α
Error tipo II:	$1 - F(-t_{\alpha,n-1} + \frac{\mu_0-\mu_1}{S/\sqrt{n}}),\ \mu_1 < \mu_0$

Y finalmente para la prueba de cola superior $H_0 : \mu = \mu_0$ *vs* $H_1 : \mu > \mu_0$ se conocen los siguientes resultados.

Prueba:	$H_0 : \mu = \mu_0 \quad vs \quad H_1 : \mu > \mu_0$
Estadística de prueba:	$t = \frac{\bar{X}-\mu_0}{S/\sqrt{n}}$
Región de rechazo:	$t \geq t_{\alpha,n-1}$, (prueba de cola superior)
Error tipo I:	α
Error tipo II:	$F(t_{\alpha,n-1} + \frac{\mu_0-\mu_1}{S/\sqrt{n}}),\ \mu_1 > \mu_0.$

EJEMPLO 2.54. Se desea determinar si la aplicación de un cierto medicamento afecta la presión arterial sistólica en el ser humano. Para ello se escogen al azar diez personas, se les mide la presión arterial, después se les aplica el medicamento y una vez que éste ha tenido efecto se mide nuevamente la presión de las personas. Se calcula entonces la diferencia entre la primera medición de la presión y la segunda. Los número obtenidos fueron los siguientes:

$$2, -1, 0, -5, 3, 2, 5, -3, 0, 4.$$

Supondremos que la diferencia calculada puede modelarse mediante una variable aleatoria con distribución normal con media μ y varianza σ^2 desconocidas. Deseamos llevar a cabo la prueba de hipótesis

$$H_0 : \mu = 0 \quad vs \quad H_1 : \mu \neq 0.$$

La primera hipótesis establece que el medicamento no influye significativamente en la presión arterial de las personas. La segunda hipótesis indica que el medicamento sí afecta la presión arterial. Con los datos obtenidos podemos calcular la media y la varianza muestral

$$\begin{aligned} \bar{x} &= 0.7, \\ s^2 &= 9.7888, \end{aligned}$$

y entonces el valor de la estadística de la prueba es

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = 0.6712.$$

Para llevar a cabo la prueba tenemos que comparar este valor con $t_{\alpha/2,n-1}$. Tomaremos $\alpha = 0.1$, y de la tabla de la distribución t encontramos que $t_{\alpha/2,n-1} = 1.833$. La regla de decisión es rechazar H_0 cuando $|t| > t_{\alpha/2,n-1}$, pero ello no sucede, por lo tanto concluimos que con base en la muestra obtenida y la prueba estadística aplicada, no existen evidencias para afirmar que el medicamento afecte la presión arterial de las personas.

Prueba acerca de la diferencia entre las medias de dos distribuciones normales con varianza conocida

Sean $X_1^1, \ldots, X_{n_1}^1$ y $X_1^2, \ldots, X_{n_2}^2$ dos muestras aleatorias independientes de dos poblaciones, ambas con distribución normal, pero con distintos parámetros, la primera con media desconocida μ_1 y varianza conocida σ_1^2, y la segunda con media desconocida

μ_2 y varianza conocida σ_2^2. Observe que el tamaño de las muestras puede ser distinto. En esta sección encontraremos un criterio para llevar a cabo las siguientes pruebas de hipótesis

$$\begin{aligned} H_0 : \mu_1 - \mu_2 = \delta \quad & vs \quad H_1 : \mu_1 - \mu_2 \neq \delta, \\ \cdots \quad & vs \quad H_1 : \mu_1 - \mu_2 < \delta, \\ \cdots \quad & vs \quad H_1 : \mu_1 - \mu_2 > \delta, \end{aligned}$$

en donde δ es una constante. Mediante estas pruebas se puede decidir si las medias de las dos poblaciones normales difieren en la constante δ o en una cantidad diferente. Consideraremos primero el caso

$$H_0 : \mu_1 - \mu_2 = \delta \quad vs \quad H_1 : \mu_1 - \mu_2 \neq \delta.$$

Denotaremos por $\bar{X}_1$ a la media muestral de la primera muestra, y por $\bar{X}_2$ a la media de la segunda muestra. Sabemos que $\bar{X}_1$ tiene distribución $\mathrm{N}(\mu_1, \sigma_1^2/n_1)$ y $\bar{X}_2$ tiene distribución $\mathrm{N}(\mu_2, \sigma_2^2/n_2)$. Entonces

$$\bar{X}_1 - \bar{X}_2 \sim \mathrm{N}(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}).$$

Este es el resultado que nos llevará a encontrar una regla para decidir cuándo rechazar H_0 en favor de H_1, con base en los datos de la muestra aleatoria. Cuando H_0 es cierta, esto es, cuando $\mu_1 - \mu_2 = \delta$, tenemos que $\bar{X}_1 - \bar{X}_2$ tiene distribución $\mathrm{N}(\delta, \sigma_1^2/n_1 + \sigma_2^2/n_2)$, y por lo tanto

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0, 1).$$

La estadística Z es nuevamente una medida natural de la distancia entre $\bar{X}_1 - \bar{X}_2$ y δ. Es entonces razonable rechazar $H_0 : \mu_1 - \mu_2 = \delta$ cuando la variable Z sea grande. Es por ello que tomamos como criterio de decisión rechazar H_0 cuando $|Z| \geq k$, para cierta constante k. En una tabla de la distribución normal estándar podemos encontrar un valor $z_{\alpha/2}$ tal que $P(|Z| \geq z_{\alpha/2}) = \alpha$, en donde α es el error tipo I. Este valor $z_{\alpha/2}$ es la constante k buscada y con ello se logra que la región de rechazo sea de tamaño α. Los detalles de esta prueba se resumen en la siguiente tabla.

Prueba:	$H_0 : \mu_1 - \mu_2 = \delta \quad vs \quad H_1 : \mu_1 - \mu_2 \neq \delta$
Estadística de prueba:	$Z = (\bar{X}_1 - \bar{X}_2 - \delta)/\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
Región de rechazo:	$\lvert Z\rvert \geq z_{\alpha/2}$, (prueba de dos colas)
Error tipo I:	α
Error tipo II:	$\Phi(z_{\alpha/2} + (\delta - \delta_1)/\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}})$
	$-\Phi(-z_{\alpha/2} + (\delta - \delta_1)/\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}), \quad \delta_1 \neq \delta.$

Comprobaremos la fórmula que aparece en la tabla anterior acerca del error tipo II. Sea δ_1 cualquier número distinto de δ. Calcularemos la probabilidad del error tipo II dado que el verdadero valor de la diferencia $\mu_1 - \mu_2$ es δ_1.

$$\begin{aligned}
\beta(\delta_1) &= P(\text{“No rechazar } H_0 \text{ cuando } \mu_1 - \mu_2 = \delta_1\text{”}) \\
&= P(\,|Z| < z_{\alpha/2} \,|\, \mu_1 - \mu_2 = \delta_1\,) \\
&= P(\,|\frac{\bar{X}_1 - \bar{X}_2 - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}| < z_{\alpha/2} \,|\, \mu_1 - \mu_2 = \delta_1\,) \\
&= P(\,\delta - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \bar{X}_1 - \bar{X}_2 < \delta + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad | \,\mu_1 - \mu_2 = \delta_1\,) \\
&= P(\,-z_{\alpha/2} + \frac{\delta - \delta_1}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < \frac{\bar{X}_1 - \bar{X}_2 - \delta_1}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < z_{\alpha/2} + \frac{\delta - \delta_1}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}\,) \\
&= \Phi(z_{\alpha/2} + \frac{\delta - \delta_1}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}) - \Phi(-z_{\alpha/2} + \frac{\delta - \delta_1}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}).
\end{aligned}$$

Para la prueba

$$H_0 : \mu_1 - \mu_2 = \delta \quad vs \quad H_1 : \mu_1 - \mu_2 < \delta,$$

los detalles se muestran en el siguiente cuadro.

Prueba:	$H_0 : \mu_1 - \mu_2 = \delta \quad vs \quad H_1 : \mu_1 - \mu_2 < \delta$
Estadística de prueba:	$Z = (\bar{X}_1 - \bar{X}_2 - \delta)/\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
Región de rechazo:	$Z \leq -z_\alpha$, (prueba de cola inferior)
Error tipo I:	α
Error tipo II:	$1 - \Phi(z_\alpha + (\delta - \delta_1)/\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}), \quad \delta_1 < \delta.$

Y para la prueba

$$H_0 : \mu_1 - \mu_2 = \delta \quad vs \quad H_1 : \mu_1 - \mu_2 > \delta,$$

los detalles se muestran a continuación.

Prueba:	$H_0 : \mu_1 - \mu_2 = \delta \quad vs \quad H_1 : \mu_1 - \mu_2 > \delta$
Estadística de prueba:	$Z = (\bar{X}_1 - \bar{X}_2 - \delta)/\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
Región de rechazo:	$Z \geq z_\alpha$, (prueba de cola superior)
Error tipo I:	α
Error tipo II:	$\Phi(z_\alpha + (\delta - \delta_1)/\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}), \quad \delta_1 > \delta.$

EJEMPLO 2.55. En una muestra aleatoria, el tiempo promedio en el que 50 mujeres terminaron una prueba escrita fue de 30 minutos, mientras que 45 hombres terminaron la prueba en un promedio de 35 minutos. Para fines ilustrativos supondremos una varianza de 9 unidades en ambas poblaciones. ¿Hay alguna diferencia entre hombres y mujeres en el tiempo promedio real para concluir la prueba? Para contestar a esta pregunta podemos llevar a cabo la prueba de hipótesis

$$H_0 : \mu_1 - \mu_2 = 0 \quad vs \quad H_1 : \mu_1 - \mu_2 \neq 0,$$

en donde μ_1 corresponde a la media de la población de mujeres, y μ_2 a la media de la población de hombres. Con lo datos recabados la estadística de la prueba toma el valor

$$z = \frac{\bar{x}_1 - \bar{x}_2 - \delta}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = -8.11\,.$$

Con $\alpha = 0.10$ se tiene que $z_{\alpha/2} = 1.65$. Entonces $|z| \geq z_{\alpha/2}$ y por lo tanto se rechaza la hipótesis nula, es decir, las poblaciones de hombres y mujeres muestran tiempos promedios diferentes para terminar la prueba escrita.

Prueba acerca de la varianza de una distribución normal

Consideremos nuevamente una muestra aleatoria $X_1, \ldots, X_n$ proveniente de n observaciones de una variable aleatoria con distribución normal con media y varianza desconocidas. Nos interesa ahora encontrar un mecanismo para probar la hipótesis de que la varianza σ^2 es igual a un valor especificado σ_0^2 o no lo es. Es decir, las hipótesis que nos interesa contrastar son

$$\begin{aligned} H_0 : \sigma^2 = \sigma_0^2 \quad & vs \quad H_1 : \sigma^2 \neq \sigma_0^2, \\ \cdots \quad & vs \quad H_1 : \sigma^2 < \sigma_0^2, \\ \cdots \quad & vs \quad H_1 : \sigma^2 > \sigma_0^2. \end{aligned}$$

Un mecanismo para encontrar una regla de decisión para estas pruebas hace uso del resultado teórico que establece que la variable aleatoria

$$\chi_0^2 := \frac{(n-1)S^2}{\sigma_0^2}$$

tiene distribución $\chi^2(n-1)$ cuando la varianza desconocida σ^2 es efectivamente σ_0^2. Como antes, el término S^2 denota la varianza muestral, es decir,

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

Para la prueba con hipótesis alternativa $H_1 : \sigma^2 \neq \sigma_0^2$, se rechaza la hipótesis nula H_0 cuando $\chi_0^2 > \chi^2_{\alpha/2,n-1}$ o bien $\chi_0^2 < \chi^2_{1-\alpha/2,n-1}$, en donde $\chi^2_{\alpha/2,n-1}$ es el número real tal que la distribución ji-cuadrada con $n-1$ grados de libertad acumula a la derecha probabilidad $\alpha/2$. Análogamente, la probabilidad a la derecha del número $\chi^2_{1-\alpha/2,n-1}$ es $1-\alpha/2$. Véase la Figura 2.20 para una representación gráfica de estas cantidades, así como de la región de rechazo de esta prueba. En la página 217 aparece una tabla que muestra las cantidades $\chi^2_{\alpha,n}$ para algunos valores de los parámetros α y n. Los detalles de esta prueba se encuentran en el recuadro siguiente:

Prueba:	$H_0 : \sigma^2 = \sigma_0^2 \quad vs \quad H_1 : \sigma^2 \neq \sigma_0^2$
Estadística de prueba:	$\chi_0^2 = (n-1)S^2/\sigma_0^2$
Región de rechazo:	$\chi_0^2 > \chi^2_{\alpha/2,n-1}$ o $\chi_0^2 < \chi^2_{1-\alpha/2,n-1}$
Error tipo I:	α
Error tipo II:	$F(\chi^2_{\alpha/2,n-1}\cdot\sigma_0^2/\sigma_1^2) - F(\chi^2_{1-\alpha/2,n-1}\cdot\sigma_0^2/\sigma_1^2)$ para $\sigma_1^2 \neq \sigma_0^2$, en donde F es la función de distribución de una v.a. $\chi^2(n-1)$.

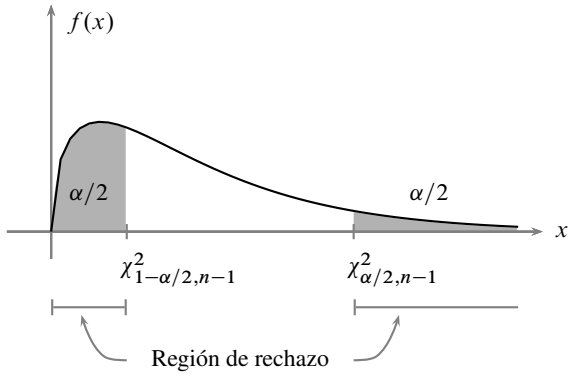


FIGURA 2.20

Como antes, el error tipo I lo establece la persona que lleva a cabo la prueba y es de magnitud α en las fórmulas anteriores. Comprobaremos ahora la fórmula del error tipo II. Sea σ_1^2 cualquier número positivo distinto de σ_0^2. Calcularemos la probabilidad

del error tipo II dado que el verdadero valor de la varianza es σ_1^2.

$$\begin{aligned}
\beta(\sigma_1^2) &= P(\text{“No rechazar } H_0\text{”} \,|\, \sigma^2 = \sigma_1^2) \\
&= P(\chi^2_{1-\alpha/2,n-1} < \frac{(n-1)S^2}{\sigma_0^2} < \chi^2_{\alpha/2,n-1} \,|\, \sigma^2 = \sigma_1^2) \\
&= P(\chi^2_{1-\alpha/2,n-1} \cdot \frac{\sigma_0^2}{\sigma_1^2} < \frac{(n-1)S^2}{\sigma_1^2} < \chi^2_{\alpha/2,n-1} \cdot \frac{\sigma_0^2}{\sigma_1^2} \,|\, \sigma^2 = \sigma_1^2) \\
&= F(\chi^2_{\alpha/2,n-1} \cdot \sigma_0^2/\sigma_1^2) - F(\chi^2_{1-\alpha/2,n-1} \cdot \sigma_0^2/\sigma_1^2),
\end{aligned}$$

en donde F es la función de distribución χ^2 con $n-1$ grados de libertad. La prueba que tiene como hipótesis alternativa $H_1 : \sigma^2 < \sigma_0^2$ tiene las siguientes características.

Prueba:	$H_0 : \sigma^2 = \sigma_0^2 \quad vs \quad H_1 : \sigma^2 < \sigma_0^2$
Estadística de prueba:	$\chi_0^2 = (n-1)S^2/\sigma_0^2$
Región de rechazo:	$\chi_0^2 < \chi^2_{1-\alpha,n-1}$
Error tipo I:	α
Error tipo II:	$1 - F(\chi^2_{1-\alpha,n-1} \cdot \sigma_0^2/\sigma_1^2)$ para $\sigma_1^2 < \sigma_0^2$, en donde F es la función de distribución de una v.a. $\chi^2(n-1)$.

Y finalmente para la prueba con hipótesis alternativa $H_1 : \sigma^2 > \sigma_0^2$ se tienen los siguientes resultados análogos.

Prueba:	$H_0 : \sigma^2 = \sigma_0^2 \quad vs \quad H_1 : \sigma^2 > \sigma_0^2$
Estadística de prueba:	$\chi_0^2 = (n-1)S^2/\sigma_0^2$
Región de rechazo:	$\chi_0^2 > \chi^2_{\alpha,n-1}$
Error tipo I:	α
Error tipo II:	$F(\chi^2_{\alpha,n-1} \cdot \sigma_0^2/\sigma_1^2)$ para $\sigma_1^2 > \sigma_0^2$, en donde F es la función de distribución de una v.a. $\chi^2(n-1)$.

Con esto concluimos nuestra breve exposición sobre pruebas de hipótesis. Existen muchas otras pruebas para rechazar o no rechazar muy diversas hipótesis estadísticas que el lector interesado puede localizar en textos como [4] y [13] o en la literatura especializada en el área de interés.

Apéndice A

Ejercicios

Espacio muestral y eventos

1. Determine un espacio muestral para cada uno de los siguientes experimentos aleatorios.
   a) Lanzar una moneda tres veces.
   b) Lanzar a un mismo tiempo tres monedas indistinguibles.
   c) Escoger un número real al azar dentro del intervalo $[-1, 1]$ y después elevarlo al cuadrado.
   d) Registrar el número de llamadas telefónicas que llegan a un número de emergencias en un minuto dado.
   e) Colocar al azar dos bolas distinguibles en cuatro celdas numeradas.
   f) Colocar al azar dos bolas indistinguibles en cuatro celdas numeradas.
2. Encuentre un espacio muestral para el experimento aleatorio de observar el marcador final de un juego de fútbol soccer. ¿Es un espacio muestral finito o infinito?
3. Suponga que se tiene en operación una sala de cómputo con 100 computadoras. Encuentre un espacio muestral para el experimento de observar la configuración de máquinas, desde el punto de vista de uso o no uso del equipo, en un momento cualquiera del día.
4. Un experimento aleatorio consiste en lanzar un dado hasta que se obtiene un "6". Proponga dos espacios muestrales para este experimento.
5. Determine un experimento aleatorio para los siguientes espacios muestrales.
   a) $\Omega = \{0, 1, 2, \ldots\}$.
   b) $\Omega = [0, \infty)$.
   c) $\Omega = \{0, 1\}$.

Operaciones con conjuntos

6. Demuestre que si $A \subseteq B$, entonces $B^c \subseteq A^c$.
7. Demuestre que si $A = A \cap B$, entonces $A \subseteq B$.
8. Demuestre que $B - A \subseteq A^c$.
9. Demuestre las leyes distributivas:

a) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
b) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

10. Demuestre las leyes de De Morgan:
    a) $(A \cup B)^c = A^c \cap B^c$.
    b) $(A \cap B)^c = A^c \cup B^c$.
11. Enuncie y demuestre las leyes de De Morgan para tres conjuntos.
    Sugerencia: use el mismo resultado para dos conjuntos.
12. Demuestre que:
    a) $A = (A \cap B) \cup (A \cap B^c)$.
    b) $A \cap B = A \cap (B \cup A^c)$.
    c) $A \cup B = A \cup (B \cap A^c)$.
13. Demuestre que
    a) $A - B = A - (A \cap B)$.
    b) $A - B = (A \cup B) - B$.
14. Demuestre que
    a) $A \cap (B - C) = (A \cap B) - (A \cap C)$.
    b) $(A - C) \cap (B - C) = (A \cap B) - C$.
15. Demuestre que las siguientes dos definiciones de la operación diferencia simétrica son equivalentes.
    a) $A \triangle B = (A - B) \cup (B - A)$.
    b) $A \triangle B = (A \cup B) - (A \cap B)$.
16. Compruebe gráficamente las siguientes propiedades básicas de la diferencia simétrica.
    a) $A \triangle (B \triangle C) = (A \triangle B) \triangle C$.
    b) $A \triangle \emptyset = A$.
    c) $A \triangle \Omega = A^c$.
    d) $A \triangle A = \emptyset$.
    e) $A \triangle A^c = \Omega$.
17. Considere los subconjuntos de números reales $A = \{x \in \mathbb{R} : x^2 \leq 1\}$ y $B = \{x \in \mathbb{R} : x > 0\}$. Encuentre y represente en el eje real los conjuntos: $A \cup B$, $A \cap B$, $A - B$ y $B - A$.
18. Sean $A = \{x \in \mathbb{R} : |x - 2| \leq 4\}$ y $B = \{x \in \mathbb{R} : |x - 1| > 2\}$. Muestre gráficamente los conjuntos A, B, A^c, B^c, $A \cap B$, $A \cup B$, $A - B$, $B - A$ y $A \triangle B$.
19. Sean $A = \{x \in \mathbb{R} : |x + 1| \leq 2\}$ y $B = \{x \in \mathbb{R} : x^2 > 2\}$. Muestre gráficamente los conjuntos A, B, A^c, B^c, $A \cap B$, $A \cup B$, $A - B$, $B - A$ y $A \triangle B$.
20. Sean $A = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$ y $B = \{(x, y) \in \mathbb{R}^2 : y > x^2\}$. Muestre gráficamente los conjuntos A^c, B^c, $A \cap B$, $A \cup B$, $A - B$, $B - A$ y $A \triangle B$.

21. Sean $A = \{(x, y) \in \mathbb{R}^2 : 2x^2 + 3y^2 \leq 6\}$ y $B = \{(x, y) \in \mathbb{R}^2 : y > -x^2\}$. Muestre gráficamente los conjuntos A^c, B^c, $A \cap B$, $A \cup B$, $A - B$, $B - A$ y $A \triangle B$.
22. Sean $A = \{(x, y) \in \mathbb{R}^2 : |x| + |y| \leq 1\}$ y $B = \{(x, y) \in \mathbb{R}^2 : y \geq x\}$. Encuentre el área del conjunto $A \cap B$.
23. *Ellos*. Si el 70 % de los hombres son feos, el 60 % son fuertes, y el 50 % son formales, ¿cuál es el porcentaje mínimo y máximo posibles de hombres que son feos, fuertes y formales?
24. *Ellas*. Si el 90 % de las mujeres son bonitas, el 80 % son sensibles, y el 70 % son sinceras, ¿cuál es el porcentaje mínimo y máximo posibles de mujeres que son bonitas, sensibles y sinceras?
25. Una compañía constructora está llevando a cabo tres proyectos de manera independiente. Sean A, B y C los eventos de que los proyectos 1, 2 y 3 sean completados a tiempo, respectivamente. Describa los siguientes eventos en términos de A, B y C.
    a) Los tres proyectos son terminados a tiempo.
    b) Sólo dos proyectos son terminados a tiempo.
    c) Sólo uno de los proyectos es terminado a tiempo.
    d) Ningún proyecto es terminado a tiempo.
    e) Sólo los proyectos A y C son terminados a tiempo.

Conjuntos ajenos

26. Compruebe que los conjuntos A y $B - A$ son siempre ajenos.
27. Demuestre que si $A \cap B = \emptyset$, entonces $A \subseteq B^c$ y $B \subseteq A^c$.
28. Suponga que A es el conjunto de raíces de la ecuación cuadrática $x^2 - x - 2 = 0$ y B es el intervalo $[0, 3]$. ¿Son los conjuntos A y B ajenos?
29. Sean A y B dos conjuntos cualesquiera. Compruebe que la colección de conjuntos: $A \cap B$, $A^c \cap B$, $A \cap B^c$ y $A^c \cap B^c$ son ajenos dos a dos. Sugerencia: dibuje un diagrama de Venn.
30. ¿Son los conjuntos $\{(x, y) \in \mathbb{R}^2 : y = e^{-x}\}$ y $\{(x, y) \in \mathbb{R}^2 : y = 1 - x\}$ ajenos? En caso negativo, ¿cuántos puntos en común tienen?
31. Sean A y B ajenos, y sea $A_1 \subseteq A$. Compruebe que A_1 y B son ajenos. Intuitivamente la afirmación es evidente, el reto consiste en encontrar una justificación matemática de ello.
32. ¿Cuántas igualdades son necesarias verificar en general para comprobar que n conjuntos son ajenos dos a dos?

Conjunto potencia

33. ¿Quién es $2^{\emptyset}$?
34. Sea $\Omega = \{a\}$. Escriba explícitamente al conjunto 2^{Ω}.
35. Demuestre que si $A \subseteq B$, entonces $2^A \subseteq 2^B$.

36. Suponga $A \subseteq B$. ¿Tiene sentido la identidad $2^A\, 2^{B-A} = 2^B$?

Producto Cartesiano

37. Sean $A = \{a_1, a_2\}$, $B = \{b_1, b_2, b_3\}$ y $C = \{c_1, c_2\}$. Escriba explícitamente a los doce elementos del producto Cartesiano $A \times B \times C$.
38. Sea $\Omega = \{a, b, c\}$. Escriba explícitamente al conjunto $\Omega \times \Omega$.
39. Sea $A = \{0, 1\}$. ¿Cuántos elementos tiene el conjunto $\underbrace{A \times \cdots \times A}_{n}$?

Probabilidad

40. Compruebe que la definición de probabilidad clásica cumple con los tres axiomas de la probabilidad.
41. Compruebe que la definición de probabilidad frecuentista cumple con los tres axiomas de la probabilidad.
42. Sean P_1 y P_2 dos medidas de probabilidad definidas sobre la misma clase de subconjuntos de Ω. Sea α un número en el intervalo $[0, 1]$. Demuestre que $P = \alpha P_1 + (1 - \alpha) P_2$ satisface los tres axiomas de la probabilidad.
43. Demuestre que $P(\emptyset) = 0$, sin usar $P(\Omega) = 1$.
44. Demuestre que $P(B - A) = P(B) - P(A \cap B)$, para cualesquiera eventos A y B.
45. Demuestre que $0 \leq P(A \cap B) \leq P(A) \leq P(A \cup B) \leq P(A) + P(B) \leq 2$.
46. Demuestre que $P(A \cap B) \geq P(A) + P(B) - 1$.
47. Sean A y B eventos ajenos tales que $P(A) = 0.3$ y $P(B) = 0.2$. Encuentre
    a) $P(A \cup B)$.
    b) $P(A^c)$.
    c) $P(A^c \cap B)$.
    d) $P(A \cap B^c)$.
    e) $P(A \triangle B)$.
48. Diga falso o verdadero. Justifique su respuesta en cada caso.
    a) Si $P(A) = 0$, entonces $P(A \cap B) = 0$.
    b) Si $P(A) = P(B)$, entonces $A = B$.
    c) Si $P(A) \leq P(B)$, entonces $A \subseteq B$.
49. Diga falso o verdadero. Justifique su respuesta en cada caso.
    a) Si $P(A) > 0$, entonces $P(A \cup B) > 0$.
    b) Si $P(A) > 0$, entonces $P(A \cap B) > 0$.
    c) Si $P(A) > 1/2$ y $P(B) > 1/2$, entonces $P(A \cap B) > 0$.
    d) Si $P(A) > 0$, entonces $P(A^c) > 0$.
50. Diga falso o verdadero. Justifique su respuesta en cada caso.
    a) $P(B - A) = P(B) - P(A)$.
    b) $P(A \cap B) = P(A)P(B)$.
    c) Si $P(A) > 1/2$, entonces $P(A^c) < 1/2$.

d) Si $P(A^c) > 0$, entonces $P(A) > 0$.
e) Si $A \cap B = \emptyset$, entonces $P(A) = P(A \cap B^c)$.

51. La probabilidad de un cierto evento es tal que duplica la probabilidad de su complemento. ¿Cuánto vale la probabilidad de este evento?
52. Demuestre que $P(A \triangle B) = P(A) + P(B) - 2P(A \cap B)$. Esta es la probabilidad de que suceda exactamente uno de los eventos A y B. Recuerde que la operación $A \triangle B$ es la diferencia simétrica entre A y B, y se define como $(A \cup B) - (A \cap B)$.
53. Demuestre que $P(A^c \cap B^c) = 1 - P(A) - P(B) + P(A \cap B)$.

Análisis combinatorio

54. ¿De cuántas formas distintas pueden seis personas formarse en una fila lineal?
55. Diez personas votarán por uno de tres candidatos.
    a) ¿De cuántas formas distintas pueden los votos ser registrados uno después de otro?
    b) ¿De cuántas formas distintas pueden los votos ser registrados en conjunto?
56. En un partido de fútbol hubo 4 goles en el marcador final. ¿Cuántas historias distintas de marcadores llevan a un marcador final donde hay 4 goles?
57. ¿De cuántas formas distintas pueden dos equipos de fútbol terminar en la clasificación general de un torneo en donde compiten 20 equipos?
58. Una enciclopedia de 5 volúmenes es colocada en un librero de modo aleatorio. Demuestre que la probabilidad de que los volúmenes queden colocados apropiadamente de derecha a izquierda o de izquierda a derecha es $1/60$.
59. ¿Cuántas diagonales se pueden trazar en un polígono convexo de n lados? No se considera que los lados sean diagonales.
60. ¿De cuántas maneras diferentes pueden clasificarse los tres primeros lugares de una carrera de 15 corredores ?
61. ¿Cuántos enteros positivos de a lo mas cinco dígitos son divisibles por 2? ¿Cuántos hay que empiecen con el dígito 1?
62. ¿Cuántos equipos de 3 personas se pueden formar de un grupo de 5 personas?
63. Al final de una reunión de 10 personas, éstas se despiden con un saludo de mano. ¿Cuántos saludos ha habido en total?
64. *Cuatro parejas.* Al final de una cena, cuatro parejas (cada una de ellas compuesta por un hombre y una mujer) se despiden de la siguiente forma: si se despiden mujer-mujer o mujer-hombre, el despido es a través de un beso. Entre dos hombres el despido es a través de un abrazo. ¿Cuántos despidos de cada tipo ha habido? Por supuesto, las dos personas que son pareja no se despiden.
65. La novela Rayuela del escritor argentino Julio Cortázar contiene 56 capítulos que aparentemente pueden ser leídos en cualquier orden. ¿De cuántas formas distintas puede ser leída esta novela?

66. En un popular juego de lotería se pide adivinar los seis números que serán escogidos al azar y sin orden dentro del conjunto $\{1, 2, \ldots, 49\}$.
    a) ¿Cuál es el total de arreglos con los cuales un jugador puede participar en este juego?
    b) Si se establece que un jugador obtiene un segundo lugar si acierta únicamente a cinco de los seis números seleccionados, ¿cuántos arreglos para segundos lugares puede haber para un arreglo dado de seis números?
67. Un cierto partido de fútbol terminó con un marcador de 4-4. ¿De cuántas formas distintas pudieron los equipos alternarse en anotar para llegar a tal resultado? ¿Cree usted que todos estos resultados son igualmente probables?
68. *Funciones.*
    a) ¿Cuántas distintas funciones existen definidas en $\{1, 2, \ldots, k\}$ y con valores en $\{1, 2, \ldots, n\}$?
    b) ¿Cuántas de estas funciones son constantes?
    c) ¿Cuántas de estas funciones son sobre? ($k \geq n$)
    d) ¿Cuántas de ellas son inyectivas? ($k \leq n$)
    e) ¿Cuántas de ellas son uno a uno?
    f) ¿Cuántas de ellas son invertibles?
69. Demuestre que:
    a) $\binom{n}{k} = \frac{n-k+1}{k}\binom{n}{k-1}$.
    b) $\binom{n}{k} = \frac{n}{n-k}\binom{n-1}{k}$.
    c) $\binom{n-1}{k} = \frac{k+1}{n}\binom{n}{k+1}$.
    d) $\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1}$. (Triángulo de Pascal)
70. ¿De cuántas formas distintas pueden cinco candidatos terminar en una elección? Suponga que es posible observar empates múltiples.
71. Debido a un error, 50 tornillos defectuosos fueron mezclados con 200 tornillos en buen estado. Si se venden 20 tornillos tomados al azar, ¿cuál es la probabilidad de que k de ellos sean defectuosos? ($0 \leq k \leq 20$).
72. ¿Cuántos números binarios diferentes se pueden obtener al usar los siete dígitos 1010101?
73. ¿Cuántas "palabras" diferentes se pueden formar con todos los caracteres (incluyendo repeticiones) de la palabra AMAR?
74. *Cumpleaños.* Calcule la probabilidad de que en un conjunto de n personas, al menos dos de ellas tengan la misma fecha de cumpleaños. Sugerencia: considere el complemento del evento de interés.

75. Encuentre el total de números enteros de cuatro dígitos sin repetición, tomados del conjunto $\{0, 1, 2, \ldots, 9\}$ de tal manera que ningún número empiece con 0. ¿Cuántos de ellos son pares y cuántos son impares?
76. En una clase de 25 estudiantes hay 15 mujeres y 10 hombres.
    a) ¿De cuántas formas puede seleccionarse un comité de tres hombres y tres mujeres?
    b) ¿De cuántas formas puede seleccionarse un comité de seis estudiantes?
    c) ¿De cuántas formas puede seleccionarse un comité de seis estudiantes todos del mismo sexo?
77. Demuestre que $\#(2^{\Omega}) = 2^{\#\Omega}$. Sugerencia: Use el método de inducción sobre el valor de n, definiendo n como la cardinalidad de Ω.
78. Demuestre que el número máximo de regiones en las que n líneas rectas dividen un plano es $1 + n(n + 1)/2$.
79. ¿Cuántos divisores diferentes tiene el número $5^3 \cdot 7^4$?
80. Se asignan 40 problemas para un examen de probabilidad. El examen consiste de 4 problemas escogidos al azar, y cada problema tiene un peso de 25 puntos. Si un alumno resuelve únicamente 20 de los problemas asignados, ¿cuál es la probabilidad de que el alumno obtenga
    a) cero puntos?
    b) 25 puntos?
    c) 50 puntos?
    d) 75 puntos?
    e) 100 puntos?
81. ¿Cuántos subconjuntos podemos obtener de un conjunto de n elementos? Escriba explícitamente todos los subconjuntos del conjunto $\{a, b, c, d\}$.
82. Sea E el conjunto de vectores de $\mathbb{R}^n$ para los cuales alguna de sus coordenadas es cero. ¿Cuántas regiones conexas tiene el conjunto $\mathbb{R}^n - E$? Sugerencia: analice los casos $n = 1, 2, 3$.
83. ¿Cuántas configuraciones diferentes se pueden obtener en un tablero de ajedrez después de
    a) la primera jugada del primer jugador?
    b) la primera jugada del segundo jugador?
84. *Teorema del binomio.* Use el método de inducción para demostrar que para cualesquiera números reales a y b, y para cualquier entero $n \geq 1$,

$$(a + b)^n = \sum_{k=0}^{n} \binom{n}{k} a^{n-k}\, b^k.$$

85. En una sala de cómputo hay seis computadoras numeradas del 1 al 6. ¿De cuántas formas distintas pueden las seis computadoras estar siendo usadas o no usadas?
86. ¿De cuántas formas posibles se puede ordenar el conjunto $\{1, 2, \ldots, 2n + 1\}$ de tal forma que cada número impar ocupe una posición impar?

87. ¿De cuántas formas posibles se puede ordenar el conjunto $\{1, 2, \ldots, 2n\}$ de tal forma que cada número par ocupe una posición par?
88. Considere el diagrama de la Figura A.1. ¿Cuántas trayectorias existen que llevan del punto A al punto B sin retroceder hacia abajo o a la izquierda?

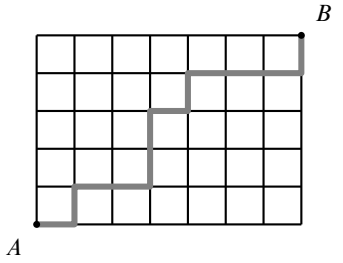


FIGURA A.1

89. ¿Cuántas "palabras" diferentes podemos obtener usando todas las letras (incluyendo repeticiones) de la palabra "manzana"?
90. ¿Cuántas "palabras" diferentes podemos obtener usando todas las sílabas (incluyendo repeticiones) de la palabra "cucurrucucú"?
91. Sea $n \geq 1$ un entero. ¿Cuántas soluciones enteras no negativas tiene la ecuación
    a) $x_1 + x_2 + \cdots + x_k = n$?
    b) $x_1 + x_2 + \cdots + x_k \leq n$?
    c) $x_1 + x_2 + \cdots + x_k \geq n$?
92. Sean $n \geq k$ dos enteros positivos. ¿Cuántos vectores con entradas enteras no negativas $(x_1, x_2, \ldots, x_k)$ satisfacen las restricciones
$$1 \leq x_1 < x_2 < \cdots < x_k \leq n\,?$$
93. Desarrolle la expresión $(a + b + c)^3$ usando la fórmula de coeficientes multinomiales (1) en la página 20, y después compruebe la fórmula directamente multiplicando el trinomio por sí mismo.
94. Demuestre que para cualesquiera enteros $n, m \geq 1$,
$$\sum \binom{n}{k_1 \cdots k_m} = m^n,$$
en donde la suma se efectúa sobre todos los posibles valores enteros no negativos de $k_1, k_2, \ldots, k_m$, tales que $k_1 + k_2 + \cdots + k_m = n$.

Probabilidad condicional e independencia

95. Sean A y B dos eventos independientes. Demuestre que
    a) A^c y B son independientes.

b) A y B^c son independientes.
c) A^c y B^c son independientes.

96. Sean A y B eventos independientes tales que $P(A) = 0.1$ y $P(B) = 0.5$. Encuentre

a) $P(A \cap B)$.
b) $P(A^c \cap B)$.
c) $P(A \cap B^c)$.
d) $P(A^c \cap B^c)$.
e) $P(A \cup B)$.
f) $P(A^c \cup B)$.
g) $P(A \cup B^c)$.
h) $P(A^c \cup B^c)$.

97. Suponga que $P(A) = 0.5$ y $P(A \cup B) = 0.6$. Encuentre $P(B)$ cuando
a) A y B son disjuntos.
b) A y B son independientes.
c) $P(A \mid B) = 0.4$.

98. Sea P una medida de probabilidad y B un evento con probabilidad positiva. Demuestre que la probabilidad condicional $P(\cdot \mid B)$ satisface los tres axiomas de la probabilidad, es decir, cumple
a) $P(A \mid B) \geq 0$.
b) $P(\Omega \mid B) = 1$.
c) $P(A_1 \cup A_2 \mid B) = P(A_1 \mid B) + P(A_2 \mid B)$ cuando $A_1 \cap A_2 = \emptyset$.

99. Suponga que $P(B) = P(A \mid B) = P(C \mid A \cap B) = p$. Demuestre que

$$P(A \cap B \cap C) = p^3.$$

100. Diga falso o verdadero. Justifique su respuesta en cada caso.
a) El número $P(A \mid B)$ nunca es cero.
b) $P(A \mid B) = P(B \mid A)$.
c) $P(A \mid B) \leq P(A)$.

101. Diga falso o verdadero. Justifique su respuesta en cada caso.
a) $P(A \mid B) + P(A^c \mid B) = 1$.
b) $P(A \mid B) + P(A \mid B^c) = P(A)$.
c) $P(A \mid A \cap B) = P(B \mid A \cap B) = 1$.

102. Diga falso o verdadero. Justifique su respuesta en cada caso.
a) $P(\Omega \mid B) = 1$.
b) $P(\emptyset \mid B) = 0$.
c) $P(A \mid A) = P(A)$.

103. Diga falso o verdadero. Justifique su respuesta en cada caso.
a) $P(A \mid B_1 \cup B_2) = P(A \mid B_1) + P(A \mid B_2)$ cuando $B_1 \cap B_2 = \emptyset$.
b) $P(A_1 \cup A_2 \mid B) = P(A_1 \mid B) + P(A_2 \mid B)$ cuando $A_1 \cap A_2 = \emptyset$.
c) $P(A \mid B_1 \cap B_2) = P(A \mid B_1)\, P(A \mid B_2)$.
d) $P(A_1 \cap A_2 \mid B) = P(A_1 \mid B)\, P(A_2 \mid B)$.

104. Demuestre que:
a) el conjunto $\emptyset$ es independiente consigo mismo.

b) el conjunto Ω es independiente consigo mismo.

105. Diga falso o verdadero. Justifique su respuesta en cada caso.
   a) A, B independientes $\Rightarrow$ A, B ajenos.
   b) A, B ajenos $\Rightarrow$ A, B independientes.
106. Sea A un evento cualquiera. Demuestre que:
   a) A y $\emptyset$ son eventos independientes.
   b) A y Ω son eventos independientes.
107. Sean A y B eventos independientes. Demuestre que

$$P(A \cup B) = 1 - P(A^c)P(B^c).$$

108. Diga falso o verdadero. Justifique su respuesta en cada caso.
   a) A, B independientes $\Rightarrow$ B, A independientes.
   b) Para cualquier evento A, A es independiente con A.
   c) A, B independientes y B, C independientes $\Rightarrow$ A, C independientes.
109. Sean A y B eventos independientes tales que $P(A) = p_1$ y $P(B) = p_2$. Calcule la probabilidad de que no ocurra ninguno de estos eventos.
110. Sean A y B eventos independientes y ajenos. Demuestre que forzosamente alguno de estos eventos tiene probabilidad cero.
111. Considere un experimento aleatorio con espacio muestral $\Omega = (0, 1)$. Definimos la probabilidad de un intervalo $(a, b) \subseteq (0, 1)$ como $P(a, b) = b - a$. Encuentre eventos A y B tales que
   a) sean independientes y ajenos.
   b) sean independientes pero no ajenos.
   c) sean ajenos pero no independientes.
   d) sean no ajenos y no independientes.
112. ¿Cuántas igualdades son necesarias verificar para demostrar que n eventos son independientes? Justifique su respuesta.
113. *Regla del producto.* Sean $A_1, \ldots, A_n$ eventos tales que la probabilidad del evento $A_1 \cap \cdots \cap A_{n-1}$ es estrictamente positiva. Compruebe que

$$\begin{aligned} P(A_1 \cap \cdots \cap A_n) &= P(A_1)\, P(A_2 \mid A_1)\, P(A_3 \mid A_1 \cap A_2) \cdots \\ &\quad \cdots P(A_n \mid A_1 \cap \cdots \cap A_{n-1}). \end{aligned}$$

114. *La urna de Polya*. El siguiente experimento se conoce como la urna de Polya. Suponga que en una urna se tienen b bolas blancas y r bolas rojas. Un experimento aleatorio consiste en seleccionar una bola al azar y regresarla a la urna junto con c bolas del mismo color. Use la regla del producto para calcular la probabilidad de obtener bolas rojas en las tres primeras extracciones.

Teorema de probabilidad total

115. *La urna de Polya*. Suponga que en una urna se tienen b bolas blancas y r bolas rojas. Un experimento aleatorio consiste en seleccionar una bola al azar

y regresarla a la urna junto con c bolas del mismo color. Sea R_n el evento "Se selecciona una bola roja en la n-ésima extracción". Compruebe que para cada $n \geq 1$,

$$P(R_n) = \frac{r}{r+b}.$$

116. Una persona toma al azar, con idéntica probabilidad, uno de los números $1, 2$ ó 3, y luego tira un dado equilibrado tantas veces como indica el número escogido. Después suma el resultado de los lanzamientos del dado. ¿Cuál es la probabilidad de que obtenga un total de 5?
117. Sea X el resultado de lanzar un dado equilibrado. Un juego consiste en seguir lanzando el dado hasta obtener un resultado Y tal que $Y \geq X$. Encuentre $P(Y = y)$ para $y = 0, 1, 2, 3, 4, 5, 6$.

Teorema de Bayes

118. Un dado equilibrado se lanza dos veces consecutivas y solamente se conoce que la suma de los dos resultados es 5. ¿Cuál es la probabilidad de haber obtenido el número 1 en el primer lanzamiento?
119. En una urna se encuentran b bolas de color blanco y n bolas de color negro. A un mismo tiempo se extraen k bolas al azar y resulta que todas son del mismo color. ¿Cuál es la probabilidad de que las bolas sean de color negro?
120. Considere nuevamente el juego en el que lanza un dado equilibrado obteniéndose un resultado X y después se lanza el dado tantas veces como sea necesario hasta obtener un valor Y tal que $Y \geq X$. ¿Cuál es la probabilidad del evento $X = 2$ dado que $Y = 6$?
121. Se escoge al azar un dígito binario para ser enviado a través de un canal de transmisión. Se escoge el "0" con probabilidad 0.4, y se escoge el "1" con probabilidad 0.6. El canal de comunicación es ruidoso de modo que un "0" se distorsiona en un "1" con probabilidad 0.2, y un "1" se distorsiona en un "0" con probabilidad 0.1. Encuentre la probabilidad de que
    a) se reciba un "0".
    b) se reciba un "1".
    c) se haya enviado un "0" dado que se recibió un "0".
    d) se haya enviado un "1" dado que se recibió un "1".
122. Una persona viaja de la población A a las poblaciones E o F con las poblaciones intermedias y probabilidades como se muestra en la Figura A.2. Si se denota por E al evento de pasar por la población E y de manera análoga para el resto de las poblaciones, calcule:
    a) $P(E)$.
    b) $P(F)$.
    c) $P(B \mid E)$.
    d) $P(D \mid F)$.

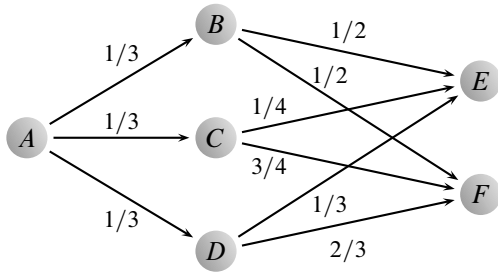


FIGURA A.2

123. En un examen de opción múltiple una pregunta escogida al azar ofrece cuatro soluciones posibles de las cuales sólo una de ellas es correcta. La probabilidad de que un estudiante sepa la respuesta correcta es 0.8. En caso de que el estudiante no conozca la respuesta, escoge cualquier opción al azar con idéntica probabilidad. Si ocurre que el estudiante ha contestado correctamente la pregunta, ¿cuál es la probabilidad de que haya sabido realmente la respuesta?

Variables aleatorias

124. Determine en cada caso si la variable aleatoria en cuestión es discreta o continua, y el conjunto de sus posibles valores.
   a) Tiempo de vida de una persona escogida al azar.
   b) Número de errores tipográficos en una página escogida al azar de un libro.
   c) Tiempo de servicio en una transacción escogida al azar realizada por una persona en un cajero automático.
   d) Monto de una reclamación por accidente automovilístico escogida al azar del conjunto de reclamaciones efectuadas a una compañía aseguradora.
125. Considere el experimento aleatorio de escoger un número ω al azar dentro del intervalo unitario $(0, 1)$. Suponga que cada resultado de este experimento se escribe en su expansión decimal como $\omega = 0.x_1x_2x_3\cdots$ Determine en los siguientes casos el conjunto de valores de la variable aleatoria definida y clasifique ésta como discreta o continua.
   a) $X(\omega) = \omega$.
   b) $X(\omega) = x_1$.
   c) $X(\omega) = 1 - \omega$.
   d) $X(w) = 0.0x_1x_2x_3\cdots$
126. Considere un experimento aleatorio con espacio muestral equiprobable $\Omega = \{1, 2, 3, 4, 5, 6\}$. Defina la variable aleatoria $X(\omega) = 2(\omega - 3)$. ¿Cuáles son los posibles valores de X? Calcule:

a) $P(X = 0)$.
b) $P(X \in \{2, 3\})$.
c) $P(X \geq 0)$.
d) $P(X < 0)$.
e) $P(X^2 = 1)$.
f) $P(2X - 4 = 0)$.
g) $P(X^2 = 4)$.
h) $P(|X - 2| \leq 4)$.

127. Considere el ejemplo del experimento aleatorio de lanzar un dardo en un tablero circular de radio uno como en la Figura 1.15 de la página 30, junto con las variables aleatorias:

$$\begin{aligned} X(x, y) &= x, \\ Y(x, y) &= y, \\ Z(x, y) &= \sqrt{x^2 + y^2}. \end{aligned}$$

Suponga que para cada región $A \subseteq \Omega$ cuya área puede ser calculada se define $P(A) = \text{Área}(A)/\text{Área}(\Omega)$. Calcule:

a) $P(X \geq 0)$.
b) $P(X < 0)$.
c) $P(Y > X)$.
d) $P(X + Y \leq 1)$.
e) $P(Z < 1/2)$.
f) $P(1/3 < Z < 1/2)$.

Funciones de densidad y de distribución

128. Encuentre el valor de la constante c para que $f(x)$ sea una función de probabilidad. Grafique esta función y calcule $P(X \in \{2, 3, 4\})$ y $P(X < 3)$ en cada caso.

a) $f(x) = \begin{cases} c\,x & \text{si } x = 1, 2, \ldots, 10, \\ 0 & \text{otro caso.} \end{cases}$

b) $f(x) = \begin{cases} c\,x^2 & \text{si } x = 1, 2, \ldots, 10, \\ 0 & \text{otro caso.} \end{cases}$

129. Encuentre el valor de la constante c para que la siguiente función sea de densidad. Grafique $f(x)$ y calcule $P(X \geq \pi)$ y $P(X \in [-\pi, 2\pi])$.

$$f(x) = \begin{cases} c\,(1 + \operatorname{sen} x) & \text{si } x \in [0, 2\pi], \\ 0 & \text{otro caso.} \end{cases}$$

130. Encuentre el valor de la constante c para que la siguiente función sea de densidad. Grafique $f(x)$ y calcule $P(X \in (1, \infty))$.

$$f(x) = c\,e^{-|x|}, \quad \text{para } -\infty < x < \infty.$$

131. Explique porqué no es posible encontrar un valor de la constante c para que la siguiente función sea de probabilidad o de densidad.

a) $f(x) = \begin{cases} c\,x & \text{si } x = -2, -1, 0, 1, 2, \\ 0 & \text{otro caso.} \end{cases}$

b) $f(x) = \begin{cases} c\ \operatorname{sen} x & \text{si } x \in [-\pi, \pi], \\ 0 & \text{otro caso.} \end{cases}$

132. Sea X una variable aleatoria discreta con función de probabilidad dada por la siguiente tabla.

x	-1	0	1
$f(x)$	0.2	0.3	0.5

Grafique $f(x)$ y calcule

a) $P(X \geq 0)$.

b) $P(X < 0)$.

c) $P(X^2 = 1)$.

133. Sea X una variable aleatoria discreta con función de probabilidad dada por la siguiente tabla.

x	-2	-1	0	2	3	5
$f(x)$	0.1	0.15	0.4	0.1	0.15	0.1

Grafique $f(x)$ y calcule la función de probabilidad de las siguientes variables aleatorias. Grafique en cada caso.

a) $Y = X^2$.

b) $Z = |X|$.

c) $W = 2X - 5$.

134. Sea X una variable aleatoria discreta con función de densidad dada por la tabla que aparece abajo. Encuentre el valor de c y grafique $f(x)$. Calcule y grafique la función de probabilidad de la variable $Y = X^2$.

x	-2	0	2
$f(x)$	0.1	c	0.1

135. Sea X una variable aleatoria continua con función de densidad

$$f(x) = \begin{cases} 1/10 & \text{si } -k \leq x \leq 4k \\ 0 & \text{otro caso.} \end{cases}$$

a) Determine el valor de la constante k y grafique $f(x)$.

b) Calcule y grafique $F(x)$.

c) Calcule $P(-1 \leq X \leq 3)$, $P(X \geq 2)$ y $P(X \leq 0)$.

d) Encuentre m tal que $P(|X - 1| \geq m) = 1/2$.

136. Sea X una variable aleatoria con función de distribución $F(x)$ como aparece abajo. ¿Es X discreta o continua? Grafique $F(x)$. Encuentre y grafique la correspondiente función de densidad $f(x)$. Calcule además $P(X = 2)$ y $P(1 < X < 2)$.

$$F(x) = \begin{cases} 0 & \text{si } x < 1, \\ 1/3 & \text{si } 1 \leq x < 2, \\ 1 & \text{si } x \geq 2. \end{cases}$$

137. Sea X una variable aleatoria con función de distribución $F(x)$ como aparece abajo. ¿Es X discreta o continua? Grafique $F(x)$. Encuentre y grafique la

correspondiente función de densidad $f(x)$. Calcule además $P(X = 1/2)$ y $P(X > 1/2)$.

$$F(x) = \begin{cases} 0 & \text{si } x < 0, \\ \sqrt{x} & \text{si } 0 \leq x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$

138. Una urna contiene cuatro bolas numeradas $1, 2, 3$ y 4. Se extraen dos bolas al azar, una a la vez y sin reemplazo. Sea X la variable aleatoria que denota la suma de los números de las bolas seleccionadas.
   a) Determine Ω.
   b) Calcule y grafique $f(x)$.
   c) Calcule y grafique $F(x)$.
   d) Calcule $P(X \geq 6)$, $P(3 < X \leq 5)$ y $P(X = 6)$.

139. Grafique las siguientes funciones y compruebe que son funciones de probabilidad.

   a) $f(x) = \begin{cases} 1/6 & \text{si } x = 0, 1, \\ 2/3 & \text{si } x = 2, \\ 0 & \text{otro caso.} \end{cases}$

   b) $f(x) = \begin{cases} \binom{4}{x} (3/4)^x \, (1/4)^{4-x} & \text{si } x = 0, 1, 2, 3, 4, \\ 0 & \text{otro caso.} \end{cases}$

140. Determine si cada una de las siguientes funciones es de densidad. Grafique la función en cada caso y justifique su respuesta.

   a) $f(x) = \begin{cases} 4x/5 & \text{si } x \in [0, 2], \\ 0 & \text{otro caso.} \end{cases}$

   b) $f(x) = \begin{cases} 2x^2/3 - 2x + 4/3 & \text{si } x \in [0, 3], \\ 0 & \text{otro caso.} \end{cases}$

141. Sea X una variable aleatoria con la función de distribución que aparece abajo. Encuentre y grafique $f(x)$. Calcule $P(0 \leq X < 10)$.

$$F(x) = \begin{cases} 1 - (1/2)^{n+1} & \text{si } n \leq x < n + 1, \\ & \text{para } n = 0, 1, 2, \ldots \\ 0 & \text{otro caso.} \end{cases}$$

142. Sea X una variable aleatoria continua con la función de densidad que aparece abajo. Encuentre el valor de la constante c y grafique la función $f(x)$. Encuentre y grafique además la función de distribución $F(x)$.

$$f(x) = \begin{cases} 2x/9 & \text{si } 0 < x < c, \\ 0 & \text{otro caso.} \end{cases}$$

143. Sea X una variable aleatoria con función de densidad

$$f(x) = \begin{cases} cxe^{-x^2} & \text{si } x > 0, \\ 0 & \text{otro caso.} \end{cases}$$

a) Determine el valor de la constante c.
b) Calcule la función de distribución.

Esperanza, varianza, momentos

144. Sea a un número fijo. Construya una variable aleatoria X tal que $E(X) = a$.
145. Calcule la esperanza de la variable aleatoria discreta X cuya función de probabilidad es:
a) $f(x) = \begin{cases} 1/3 & \text{si } x = 0, 1, \\ 1/6 & \text{si } x = 2, 3, \\ 0 & \text{otro caso.} \end{cases}$
b) $f(x) = \begin{cases} 1/4 & \text{si } x = -1, 1, \\ 1/2 & \text{si } x = 0, \\ 0 & \text{otro caso.} \end{cases}$
146. Calcule la esperanza de la variable aleatoria continua X cuya función de densidad es:
a) $f(x) = e^{-x}$, para $x > 0$.
b) $f(x) = 6x(1-x)$, para $0 < x < 1$.
147. Sea X una variable aleatoria discreta con la función de probabilidad que aparece abajo. Demuestre que $f(x)$ es efectivamente una probabilidad de densidad y que la esperanza de X no existe. Este es un ejemplo de una variable aleatoria discreta que no tiene esperanza finita.

$$f(x) = \begin{cases} \dfrac{1}{x(x+1)} & \text{si } x = 1, 2, 3, \ldots \\ 0 & \text{otro caso.} \end{cases}$$

148. Sea X una variable aleatoria continua con la función de densidad que aparece abajo. Demuestre que esta función es efectivamente una función de densidad. Compruebe además que la esperanza de X no existe. Este es un ejemplo de una variable aleatoria continua que no tiene esperanza finita. Es un caso particular de la distribución Cauchy.

$$f(x) = \frac{1}{\pi(1+x^2)}, \qquad \text{para} -\infty < x < \infty.$$

149. Diga falso o verdadero. Justifique su respuesta en cada caso.
a) La esperanza de una v.a. puede ser cero.
b) No hay dos v.a.s distintas con la misma esperanza.
c) La esperanza de una v.a. nunca es negativa.
d) La varianza de una v.a. puede ser cero.
e) La varianza de una v.a. nunca es negativa.
f) No hay dos v.a.s distintas con la misma varianza.
150. Demuestre que
a) $E(E(X)) = E(X)$.

b) $\text{Var}(\text{Var}(X)) = 0$.

151. Sea X la variable aleatoria constante c. Compruebe que:
   a) $E(X) = c$.
   b) $E(X^n) = c^n$.
   c) $\text{Var}(X) = 0$.

152. Calcule la media y varianza de la variable aleatoria X con función de probabilidad:
   a) $f(x) = \begin{cases} 1/9 & \text{si } x = 0, 1, 2, \\ 2/9 & \text{si } x = 3, 4, 5, \\ 0 & \text{otro caso.} \end{cases}$
   b) $f(x) = \begin{cases} (1/2)^{x+1} & \text{si } x = 0, 1, 2, 3, \ldots \\ 0 & \text{otro caso.} \end{cases}$

153. Diga falso o verdadero. Justifique su respuesta en cada caso.
   a) $\text{Var}(E(X)) = 0$.
   b) $E(\text{Var}(X)) = \text{Var}(X)$.

154. Sea X una variable aleatoria continua con función de densidad

$$f(x) = \frac{1}{2}e^{-|x|}, \qquad \text{para } -\infty < x < \infty.$$

   Demuestre que:
   a) $f(x)$ es efectivamente una función de densidad.
   b) $E(X) = 0$.
   c) $E(X^2) = 2$.
   d) $\text{Var}(X) = 2$.
   e) $E(X^n) = n!$ para n par.

155. Diga falso o verdadero. Justifique su respuesta en cada caso.
   a) $E(-X) = -E(X)$.
   b) $\text{Var}(-X) = -\text{Var}(X)$.
   c) $E(\text{Var}(X)) = \text{Var}(E(X))$.

156. Encuentre el error en la siguiente "demostración" de la afirmación de que la varianza de cualquier variable aleatoria es cero.

$$\begin{aligned} 0 &= \text{Var}(0) \\ &= \text{Var}(X + (-X)) \\ &= \text{Var}(X) + \text{Var}(-X) \\ &= \text{Var}(X) + \text{Var}(X) \\ &= 2\ \text{Var}(X). \end{aligned}$$

157. Sea $\theta > 0$ fijo. Encuentre el n-ésimo momento de una variable aleatoria X con función de densidad dada por

$$f(x) = \begin{cases} \theta x^{\theta-1} & \text{si } 0 < x < 1, \\ 0 & \text{otro caso.} \end{cases}$$

Cuantiles

158. Calcule los cuatro cuartiles de la variable aleatoria continua X con función de densidad
$$f(x) = \begin{cases} 2x & \text{si } 0 < x < 1, \\ 0 & \text{otro caso.} \end{cases}$$

159. Calcule los cuatro cuartiles de la variable aleatoria discreta X con función de probabilidad
$$f(x) = \begin{cases} 1/5 & \text{si } x = -2, -1, 0, 1, 2, \\ 0 & \text{otro caso.} \end{cases}$$

160. Calcule el cuantil al 75 % para una variable aleatoria discreta con función de probabilidad
$$f(x) = \begin{cases} \dfrac{1}{x(x+1)} & \text{si } x = 1, 2, 3, \ldots \\ 0 & \text{otro caso.} \end{cases}$$

161. Proporcione un ejemplo de una distribución de probabilidad discreta tal que los cuantiles al 25 % y 75 % coincidan.

Distribución uniforme discreta

162. Sea X con distribución uniforme en el conjunto $\{1, \ldots, n\}$. Demuestre que
   a) $E(X) = (n+1)/2$.
   b) $E(X^2) = (n+1)(2n+1)/6$.
   c) $\text{Var}(X) = (n^2-1)/12$.

163. Se escogen completamente al azar y de manera independiente dos números a y b dentro del conjunto $\{1, 2, \ldots, 9, 10\}$. ¿Cuál es la probabilidad de que el cociente a/b sea menor a uno?

164. Sean X y Y variables aleatorias con la misma distribución uniforme en el conjunto $\{1, 2, \ldots, 100\}$. Suponga que estas variables son independientes, es decir, los eventos $(X = x)$ y $(Y = y)$ son independientes para cualquier valor de x y y. Calcule la probabilidad del evento:
   a) $(X = Y)$.
   b) $(X \leq Y)$.
   c) $(X + Y \leq 50)$.
   d) $(X - Y \geq 10)$.

Distribución Bernoulli

165. Sea X una variable aleatoria con distribución $\text{Ber}(p)$. Demuestre que
   a) $f(x)$ es efectivamente una función de probabilidad.
   b) $E(X) = p$.
   c) $E(X^n) = p$, para $n \geq 1$.

d) $\text{Var}(X) = p(1-p)$.

166. Sean X_1 y X_2 variables aleatorias con distribución $\text{Ber}(p)$ e independientes en el sentido de que los eventos $(X_1 = x_1)$ y $(X_2 = x_2)$ son independientes para cualesquiera valores x_1 y x_2. Calcule:
   a) $P(X_1 = 0 \mid X_1 + X_2 = 1)$.
   b) $P(X_1 = 1 \mid X_1 + X_2 = 1)$.
   c) $P(X_1 + X_2 = 2 \mid X_1 = 1)$.
167. Sea X con distribución $\text{Ber}(p)$. Defina la variable aleatoria $Y = 2X - 1$.
   a) Encuentre la distribución de Y.
   b) Calcule la esperanza y varianza de Y.
   c) Encuentre el n-ésimo momento de Y.

Distribución binomial

168. Sea X una variable aleatoria con distribución $\text{bin}(n, p)$. Demuestre que:
   a) $f(x)$ es efectivamente una función de probabilidad.
   b) $E(X) = np$.
   c) $\text{Var}(X) = np(1-p)$.
169. Sea X una variable aleatoria con distribución $\text{bin}(n, p)$.
   a) Determine los valores de n y p cuando $E(X) = 4$ y $\text{Var}(X) = 2$.
   b) Determine los valores de n y p cuando $E(X) = u$ y $\text{Var}(X) = v$.
170. Sea X una variable aleatoria con distribución $\text{bin}(n, p)$. Demuestre que la variable $Y = n - X$ tiene distribución $\text{bin}(n, 1-p)$. Proporcione una explicación probabilista de este resultado.
171. Sea X con distribución $\text{bin}(n, p)$. Demuestre que para $x = 0, 1, \ldots, n-1$, se cumple la fórmula que aparece abajo. Esta expresión permite calcular las probabilidades de esta distribución de una forma iterativa.

$$P(X = x+1) = \frac{p}{(1-p)} \frac{(n-x)}{(x+1)} P(X = x).$$

172. Se lanza una moneda equilibrada 6 veces. Calcule la probabilidad de que cada cara caiga exactamente 3 veces.
173. Se lanza una moneda equilibrada $2n$ veces. Calcule la probabilidad de que ambas caras caigan el mismo número de veces.
174. Sea X una variable aleatoria con distribución $\text{bin}(n, p)$. Demuestre que

$$0 \leq \text{Var}(X) \leq E(X).$$

175. Suponiendo que es igualmente probable que nazca un hombre (H) o una mujer (M), y considerando la observación de 6 nacimientos. ¿Cuál de los siguientes tres eventos es más probable que ocurra?
   a) MHHMHM.
   b) MMMMHM.
   c) HMHMHM.

176. Un basquetbolista encesta el 95 % de sus tiros libres. Encuentre la probabilidad de que en sus siguientes diez tiros
   a) falle dos.
   b) no falle ninguno.

Distribución geométrica

177. Sea X una variable aleatoria con distribución geo(p). Demuestre que:
   a) $f(x)$ es efectivamente una función de probabilidad.
   b) $E(X) = (1-p)/p$.
   c) $\text{Var}(X) = (1-p)/p^2$.
178. *Pérdida de memoria.* Sea X una variable aleatoria con distribución geo(p). Demuestre que para cualesquiera $a, b = 0, 1, 2, \ldots$ se cumple la siguiente propiedad llamada de pérdida de memoria:

$$P(X \geq a + b \mid X \geq a) = P(X \geq b).$$

179. Sea X una variable aleatoria con distribución geo(p). Compruebe que para cualquier entero $x \geq 0$,

$$P(X \geq x) = (1-p)^x.$$

180. Considere una urna con 3 bolas negras y 5 bolas blancas. Se escoge una bola al azar, se registra su color, y después se regresa a la urna. ¿Cuántas extracciones en promedio se necesitan realizar hasta obtener una bola negra por primera vez?
181. Calcule el número promedio de intentos necesarios para obtener por primera vez un
   a) 6 en una sucesión de lanzamientos de un dado equilibrado.
   b) doble 6 en una sucesión de lanzamientos a un mismo tiempo de dos dados equilibrados indistinguibles.

Distribución Poisson

182. Sea X una variable aleatoria con distribución Poisson(λ). Demuestre que:
   a) $f(x)$ es efectivamente una función de probabilidad.
   b) $E(X) = \lambda$.
   c) $\text{Var}(X) = \lambda$.
183. Sea X una variable aleatoria con distribución Poisson(λ). Demuestre que para $x = 0, 1, 2, \ldots$ se cumple la fórmula que aparece abajo. Esta expresión permite calcular las probabilidades Poisson de una forma iterativa.

$$P(X = x+1) = \frac{\lambda}{x+1}\, P(X = x).$$

184. Sea X una variable aleatoria con distribución Poisson(λ). Demuestre que

$$P(X \in \{0, 2, 4, \ldots\}) = \frac{1}{2}\,(1 + e^{-2\lambda}).$$

185. El número de computadoras que fallan por mes en un laboratorio de cómputo tiene una distribución Poisson con un promedio mensual de $\lambda = 2$ máquinas descompuestas. El laboratorio tiene capacidad para reparar hasta dos máquinas por mes. Cuando se descomponen mas de dos máquinas, las restantes se envían fuera del laboratorio para su reparación.
   a) ¿Cuál es la probabilidad de que en un mes cualquiera sea necesario enviar máquinas fuera del laboratorio para su reparación?
   b) Responda al inciso anterior cuando se reduce la capacidad de reparación del laboratorio a una computadora por mes.
   c) ¿Cuál es el número de computadoras con falla más probable en un mes?

186. Una compañía de renta de autos estima que el número X de clientes que solicitan rentar un auto tipo deportivo en un día cualquiera puede modelarse mediante un variable aleatoria con distribución Poisson de parámetro $\lambda = 2$. Calcule la probabilidad de que en un día cualquiera
   a) ningún cliente solicite rentar un auto tipo deportivo.
   b) se renten los cuatro autos deportivos que tiene a disposición.

187. Una compañía aseguradora tiene una cartera particular de 550 pólizas contra robo de auto. La probabilidad de que cada uno de estas pólizas presente una reclamación es $p = 0.02$. Calcule la probabilidad de que un máximo de diez pólizas presenten reclamación y compare este resultado con la aproximación Poisson.

Distribución binomial negativa

188. Sea X una variable aleatoria con distribución binneg(r, p). Demuestre que:
   a) $f(x)$ es efectivamente una función de probabilidad.
   b) $E(X) = r(1-p)/p$.
   c) $\text{Var}(X) = r(1-p)/p^2$.

189. Sea X una variable aleatoria con distribución binneg(r, p). Demuestre que para $x = 1, 2, \ldots$ se cumple la fórmula que aparece abajo. Esta expresión permite calcular las probabilidades de esta distribución de una forma iterativa.

$$P(X = x) = (1-p)\,\frac{r + x - 1}{x}\,P(X = x - 1).$$

190. Se lanza de manera sucesiva una moneda equilibrada con cada uno de los lados denominado como cara o cruz.
   a) ¿Cuál es el número promedio de caras que se aparecerán hasta obtener cinco cruces?

b) ¿Cuál es el número esperado de lanzamientos que se necesitan hasta obtener cinco cruces?

Distribución hipergeométrica

191. Sea X con distribución hipergeo(N, K, n) en el caso sencillo cuando el tamaño n de la muestra es menor igual tanto al parámetro K como a $N - K$. Compruebe que $f(x)$ es efectivamente una función de probabilidad.
192. Suponga que se tiene un lote de 20 artículos de los cuales el 10 % son defectuosos y el restante 90 % se encuentran en buen estado. Se seleccionan al azar y sin reemplazo 5 artículos. Calcule la probabilidad de que:
    a) ningún artículo seleccionado sea defectuoso.
    b) dos de los artículos seleccionados sean defectuosos.
193. Cinco vehículos de carga cruzarán la frontera y uno de ellos lleva productos no permitidos. Se seleccionarán al azar dos vehículos para su inspección. ¿Cuál es la probabilidad de que se detecten los productos no permitidos?
194. Un productor desea vender un lote de 1000 artículos y sabe que 100 de ellos están defectuosos. El comprador planea tomar una muestra al azar de 20 artículos para su inspección, si encuentra por lo menos un artículo defectuoso, entonces no compra el lote. ¿Cuál es la probabilidad de que se lleve a cabo la compra?

Distribución uniforme continua

195. Sea X con distribución unif(a, b). Demuestre que:
    a) $f(x)$ es efectivamente una función de densidad.
    b) $E(X) = (a + b)/2$.
    c) $\text{Var}(X) = (b - a)^2/12$.
196. Sea X con distribución uniforme en el intervalo (a, b) tal que $E(X) = 4$ y $\text{Var}(X) = 3$. Encuentre los valores de a y b.
197. Sea X con distribución uniforme en el intervalo $(0, 1)$. Demuestre que
$$E(X^n) = \frac{1}{n+1}.$$
198. Sea X con distribución uniforme en el intervalo (a, b). Demuestre que
$$E(X^n) = \frac{b^{n+1} - a^{n+1}}{(n+1)(b-a)}.$$
199. Sea X una variable aleatoria con distribución uniforme en el intervalo $(-1, 1)$. Demuestre que
$$E(X^n) = \begin{cases} 1/(n+1) & \text{si } n \text{ es par,} \\ 0 & \text{si } n \text{ es impar.} \end{cases}$$

200. Sea X con distribución unif$(0, 1)$. Obtenga la distribución de la variable $Y = 10X - 5$.
201. Sea X con distribución unif(a, b). Demuestre que la media y la mediana de X coinciden.

Distribución exponencial

202. Sea X una variable aleatoria con distribución exp(λ). Demuestre que:
   a) $f(x)$ es efectivamente una función de densidad.
   b) $E(X) = 1/\lambda$.
   c) $\text{Var}(X) = 1/\lambda^2$.
203. *Pérdida de memoria.* Sea X una variable aleatoria con distribución exp(λ). Demuestre que para cualesquiera números $x, y \geq 0$ se cumple la igualdad que aparece abajo. La distribución exponencial es la única distribución continua que satisface esta propiedad llamada de pérdida de memoria.

$$P(X \geq x + y \mid X \geq x) = P(X \geq y).$$

204. Sea X con distribución exp(λ). Demuestre que para cualesquiera números $x, y \geq 0$,

$$F(x + y) = F(x) + F(y) - F(x)\, F(y).$$

205. Suponga que el tiempo que un usuario cualquiera permanece conectado a un servidor en una red de cómputo se puede modelar como una variable aleatoria con distribución exponencial con media igual a 10 minutos. De mil usuarios, ¿cuántos tienen un conexión superior a una hora? Calcule además la probabilidad de que un usuario cualquiera
   a) no permanezca conectado mas de 10 minutos.
   b) permanezca conectado mas de 10 minutos pero menos de una hora.
206. *Discretización.* Sea X con distribución exp(λ). Se puede discretizar esta variable aleatoria continua de la siguiente forma: sea N una variable discreta tal que para cada $n = 0, 1, 2, \ldots$ se define

$$P(N = n) = P(n \leq X \leq n + 1).$$

   Demuestre que N tiene distribución geo(p) con $p = 1 - e^{-\lambda}$.
207. Sea X con distribución exponencial. Encuentre el valor del parámetro λ si se sabe que el cuantil al 75 % es 2.

Distribución gama

208. Sea X una variable aleatoria con distribución gama(n, λ). Demuestre que:
   a) $f(x)$ es efectivamente una función de densidad.
   b) $E(X) = n/\lambda$.
   c) $\text{Var}(X) = n/\lambda^2$.
   d) $E(X^m) = \Gamma(m + n)/(\lambda^m \Gamma(n))$.

209. Demuestre las siguientes propiedades de la función gama.
   a) $\Gamma(n+1) = n\,\Gamma(n)$.
   b) $\Gamma(2) = \Gamma(1) = 1$.
   c) $\Gamma(n+1) = n!$ si n es entero.
   d) $\Gamma(1/2) = \sqrt{\pi}$.

Distribución beta

210. Sea X con distribución beta(a, b). Demuestre que:
   a) $f(x)$ es efectivamente una función de densidad.
   b) $E(X) = a/(a+b)$.
   c) $\text{Var}(X) = ab/((a+b+1)(a+b)^2)$.
211. Demuestre las siguientes propiedades de la función beta.
   a) $B(a, b) = B(b, a)$.
   b) $B(a, 1) = 1/a$.
   c) $B(1, b) = 1/b$.
   d) $B(a+1, b) = \dfrac{a}{b}\, B(a, b+1)$.
   e) $B(a+1, b) = \dfrac{a}{a+b}\, B(a, b)$.
   f) $B(a, b+1) = \dfrac{b}{a+b}\, B(a, b)$.
   g) $B(1/2, 1/2) = \pi$.
212. Sea X una variable aleatoria con distribución beta(a, b), con $a = b = 1$. Demuestre que X tiene una distribución uniforme en el intervalo $(0, 1)$.
213. *Distribución arcoseno*. Sea X con distribución beta$(1/2, 1/2)$. En este caso se dice que X tiene una distribución arcoseno
   a) Calcule y grafique $f(x)$.
   b) Demuestre que $f(x)$ es una función de densidad.
   c) Calcule $E(X)$ y $\text{Var}(X)$.
214. Sea X con distribución beta(a, b). Demuestre que:
   a) si $a > 0$ y $b = 1$, entonces la función de distribución de X toma la forma
$$F(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ x^a & \text{si } 0 < x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$
   b) si $a = 1$ y $b > 0$, entonces la función de distribución de X toma la forma
$$F(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ 1-(1-x)^b & \text{si } 0 < x < 1, \\ 1 & \text{si } x \geq 1. \end{cases}$$
215. Sea X con distribución beta(a, b). Demuestre que para cualquier entero $n \geq 0$,
$$E(X^n) = \frac{B(a+n, b)}{B(a, b)}.$$

Distribución Weibull

216. Mediante el cambio de variable $u(x) = (\lambda x)^{\alpha}$, demuestre que la función de densidad Weibull(α, λ) es efectivamente una función de densidad.
217. Nuevamente mediante el cambio de variable $u(x) = (\lambda x)^{\alpha}$, compruebe que la función de acumulación de la distribución Weibull(α, λ), es

$$F(x) = \begin{cases} 1 - e^{-(\lambda x)^{\alpha}} & \text{si } x > 0, \\ 0 & \text{otro caso.} \end{cases}$$

218. Suponga que un cierto componente eléctrico tiene un tiempo de vida útil en años que se considera aleatorio pero modelado mediante una variable aleatoria X con distribución Weibull de parámetros $\alpha = 1/2$ y $\lambda = 1$. Compruebe que la probabilidad de que el componente permanezca en buenas condiciones los primeros cinco años es 0.10697. ¿Cuánto tiempo en promedio se espera que funcione bien este componente?

Distribución normal

219. Sea X con distribución N(μ, σ^2). Demuestre que:
    a) $f(x)$ es efectivamente una función de densidad.
    b) $E(X) = \mu$.
    c) $\text{Var}(X) = \sigma^2$.
220. *Estandarización.* Sea X con distribución N(μ, σ^2). Demuestre que la variable $Z = (X - \mu)/\sigma$ tiene distribución normal estándar. Recíprocamente, demuestre que si Z tiene distribución normal estándar, y $\sigma > 0$, entonces la variable $X = \mu + \sigma Z$ tiene distribución N(μ, σ^2).
221. Sea X con distribución N$(10, 25)$. Use la tabla de la distribución normal para calcular:
    a) $P(X \geq 10)$.
    b) $P(X < 0)$.
    c) $P(0 < X \leq 10)$.
    d) $P(X \geq 20)$.
    e) $P(-20 < X \leq 10)$.
222. Sea X con distribución N$(0, 100)$. Use la tabla de la distribución normal para calcular:
    a) $P(X \leq 10)$.
    b) $P(X > 0)$.
    c) $P(0 < X \leq 40)$.
    d) $P(X \geq 30)$.
    e) $P(-10 < X \leq 10)$.
223. Use la tabla de la distribución normal para encontrar el valor de x tal que:
    a) $\Phi(x) = 0.8666$.
    b) $1 - \Phi(x) = 0.9154$.

224. Sea X una variable aleatoria con distribución $\mathrm{N}(\mu, \sigma^2)$. Demuestre que la variable $Y = aX + b$, con $a \neq 0$, también tiene una distribución normal. Encuentre los parámetros correspondientes.
225. Sea X una variable aleatoria con distribución $\mathrm{N}(\mu, \sigma^2)$. Demuestre que la variable $Y = -X$ también tiene una distribución normal. Encuentre los parámetros correspondientes.
226. Sea X una variable aleatoria con distribución $\mathrm{N}(0, 1)$. Demuestre que X^2 tiene una distribución $\chi^2(1)$.
Sugerencia: Para $x > 0$, $F_{X^2}(x) = P(X^2 \leq x) = P(-\sqrt{x} \leq X \leq \sqrt{x})$.
227. Sea X una variable aleatoria con distribución normal estándar. Encuentre la función de densidad de $Y = |X|$.
228. Encuentre el primer, segundo y tercer cuartil de la distribución normal de media $\mu = 2$ y varianza $\sigma^2 = 5$.
229. Una máquina automática despachadora de refresco está ajustada para llenar vasos de 300 ml. en promedio. Debido a cuestiones mecánicas, el llenado de los vasos no es enteramente exacto y hay pequeñas fluctuaciones. El fabricante de la máquina sabe que el llenado de los vasos se puede modelar como una variable aleatoria normal con media de 300 ml. y desviación estándar $\sigma = 10$ ml.
    a) ¿Qué fracción de los vasos serán servidos con más de 310 ml?
    b) ¿Cuál es la probabilidad de que un vaso sea servido entre 290 y 305 ml?
    c) Si el restaurant utiliza vasos de 320 ml. ¿qué porcentaje de ellos se derramarán?
    d) Si los clientes reclaman por vasos servidos con 270 ml. o menos, de mil clientes, ¿cuántos de ellos reclamarán?

Distribución ji cuadrada

230. Sea X con distribución $\chi^2(n)$. Demuestre que:
    a) $f(x)$ es efectivamente una función de densidad.
    b) $E(X) = n$.
    c) $\mathrm{Var}(X) = 2n$.
    d) $E(X^m) = 2^m\, \Gamma(\frac{n}{2} + m)/\Gamma(\frac{n}{2})$, para $m \geq 0$.
231. Compruebe que la distribución $\chi^2(n)$ se reduce a la distribución $\exp(1/2)$ cuando $n = 2$.

Distribución t

232. Sea X con distribución $t(n)$. Demuestre que:
    a) $f(x)$ es efectivamente una función de densidad.
    b) $E(X) = 0$, para $n > 1$.
    c) $\mathrm{Var}(X) = n/(n-2)$, para $n > 2$.

233. Use la tabla de la distribución t al final de este libro para aproximar el valor de la integral:
   a) $\displaystyle\int_{-\infty}^{1} \frac{dx}{1+x^2}$.
   b) $\displaystyle\int_{-\infty}^{9} \frac{dx}{(12+x^2)^2}$.

Simulación de v.a.s discretas

234. Sea X una v.a. discreta con función de probabilidad $f(x)$ como aparece en la tabla de abajo. Explique una forma de simular observaciones de X. Implemente el procedimiento propuesto en una computadora y verifique numéricamente que $P(X \leq 0) = 1/4$.

x	-1	0	1	2
$f(x)$	1/8	1/8	3/8	3/8

235. Sea X una v.a. discreta con función de probabilidad $f(x)$ como aparece en la tabla de abajo. Mediante simulaciones de valores observados para X calcule un valor aproximado para $E(\cos(\pi X))$.

x	-1	0	1
$f(x)$	1/4	1/2	1/4

236. Explique una forma de simular observaciones de una variable aleatoria con distribución geo(p).
237. Explique una forma de simular observaciones de una variable aleatoria con distribución binneg(r, p).

Simulación de v.a.s continuas

238. Sea X con distribución unif$(0, 1)$. Use algún método de simulación para obtener valores de la v.a. X^2 y estimar $E(X^2)$. Use después el teorema del estadístico inconsciente que aparece en la página 42 para verificar que $E(X^2) = 1/3$.
239. Sean X y Y variables aleatorias independientes con distribución exp(λ_1) y exp(λ_2), respectivamente, en donde $\lambda_1 = 1$ y $\lambda_2 = 2$. Genere valores al azar para X y Y y compruebe de manera aproximada que

$$P(Y > X) = \frac{\lambda_1}{\lambda_1 + \lambda_2} = \frac{1}{3}.$$

240. Sea X con distribución unif$(0, 1)$. Generando valores al azar de esta variable aleaotoria y usando la ley de los grandes números aproxime la esperanza de

$$Y = \frac{1}{\sqrt{2\pi}}\, e^{-X^2/2}.$$

Esta es una forma de aproximar la probabilidad $\Phi(1) - \Phi(0) = 0.3413$.

Variables y tipos de datos

241. Clasifique cada una de las siguiente variables de acuerdo a si es cualitativa o cuantitativa, y también de acuerdo a su posible escala de medición.
   a) Evaluación de un curso usando la siguiente escala:
   MB=Muy Bien, B=Bien, S=Suficiente, NA=No Acreditado.
   b) Número de hijos de una persona.
   c) Escolaridad de una persona.
   d) Porcentaje de artículos defectuosos producidos en una fábrica.
   e) Nivel de satisfacción de una persona al adquirir un producto.
   f) Preferencia electoral por alguna de las siguientes propuestas políticas: A, B, C, D, E, F.
   g) Uso o no uso de lentes de una persona.
   h) Principal país de visita para una persona que viaja al extranjero.
   i) Costo económico en un accidente automovilístico.

Estadística descriptiva

242. Calcule la media, moda, mediana, varianza, desviación estándar y rango del conjunto de datos $\{4, 2, 0, 9, 4, 2, -1, 1, -4, 2\}$.
243. *Transformación lineal de datos.* Sea $x_1, x_2, \ldots, x_n$ una colección de números y defina la transformación $u = ax + b$, en donde a y b son dos constantes. De este modo se obtiene la colección de datos transformados $u_1, u_2, \ldots, u_n$. Demuestre que:
   a) $\bar{u} = a\bar{x} + b$.
   b) $s^2(u_1, \ldots, u_n) = a^2\, s^2(x_1, \ldots, x_n)$, en donde $s^2(u_1, \ldots, u_n)$ denota la varianza muestral de los valores $u_1, \ldots, u_n$, con análogo significado para $s^2(x_1, \ldots, x_n)$.
244. Pregunte a diez personas sus estaturas. Registre todos estos datos y calcule la media, moda, mediana, varianza, desviación estándar y rango.
245. Mediante el uso de una calculadora o computadora genere diez números al azar dentro del intervalo unitario $(0, 1)$. Escriba estos datos y calcule la media, moda, mediana, varianza, desviación estándar y rango. Conforme el número de datos se incrementa, ¿hacia dónde tienden las cantidades calculadas?
246. Demuestre que
   a) $\displaystyle\sum_{i=1}^{n} (x_i - \bar{x}) = 0.$
   b) $\displaystyle\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2.$

c) $s^2 = \dfrac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - \dfrac{1}{n}(\sum_{i=1}^{n} x_i)^2\right)$.

247. Encuentre el valor de c que minimiza la función

$$\bar{x}(c) = \sum_{i=1}^{n} (x_i - c)^2.$$

248. Calcule la media, varianza, desviación estándar, mediana y moda aproximados del siguiente conjunto de datos agrupados. Elabore además un histograma.

Intervalo de clase	Frecuencia
$10 \leq x < 20$	4
$20 \leq x < 30$	3
$30 \leq x < 40$	6
$40 \leq x < 50$	5
$50 \leq x < 60$	5

249. Calcule la media, varianza, desviación estándar, mediana y moda aproximados del siguiente conjunto de datos agrupados. Elabore además un histograma.

Intervalo de clase	Frecuencia
$0 \leq x < 5$	12
$5 \leq x < 10$	23
$10 \leq x < 15$	10
$15 \leq x < 20$	14
$25 \leq x < 30$	6
$30 \leq x < 35$	10
$35 \leq x < 40$	5

Vectores aleatorios

250. En una urna hay cuatro bolas numeradas del 1 al 4. Se extraen al azar dos bolas, una después de la otra, sin reemplazo. Sea X el número de la primera bola extraída, y sea Y el número de la segunda bola extraída. Encuentre la función de probabilidad del vector aleatorio discreto (X, Y). A partir de esta función encuentre las funciones de probabilidad marginales de X y Y.

251. Sea (X, Y) un vector aleatorio discreto con función de probabilidad $f(x, y)$ dada por la siguiente tabla.

x\y	-1	0
1	0.3	0.5
2	0.05	0.15

a) Grafique $f(x, y)$ y demuestre que efectivamente se trata de una función de densidad.

b) Calcule y grafique las densidades marginales $f_1(x)$ y $f_2(y)$. Verifique que ambas son efectivamente funciones de densidad.

c) Calcule y grafique las distribuciones marginales $F_1(x)$ y $F_2(y)$. Verifique que ambas son efectivamente funciones de distribución.

d) ¿Son X y Y independientes?

252. Sea (X, Y) un vector aleatorio discreto con función de densidad $f(x, y)$ dada por la siguiente tabla.

x\y	0	1
1	0.2	0
2	0.7	0.1

a) Grafique $f(x, y)$ y demuestre que efectivamente se trata de una función de densidad.

b) Calcule y grafique las densidades marginales $f_1(x)$ y $f_2(y)$. Verifique que ambas son efectivamente funciones de densidad.

c) Calcule y grafique las distribuciones marginales $F_1(x)$ y $F_2(y)$. Verifique que ambas son efectivamente funciones de distribución.

d) ¿Son X y Y independientes?

253. Sea (X, Y) un vector aleatorio continuo con distribución uniforme en el cuadrado $(-1, 1) \times (-1, 1)$.

a) Grafique $f(x, y)$ y demuestre que efectivamente se trata de una función de densidad.

b) Calcule y grafique las densidades marginales $f_1(x)$ y $f_2(y)$. Verifique que ambas son efectivamente funciones de densidad.

c) Calcule y grafique las distribuciones marginales $F_1(x)$ y $F_2(y)$. Verifique que ambas son efectivamente funciones de distribución.

d) ¿Son X y Y independientes?

254. Sea (X, Y) un vector aleatorio continuo con función de densidad $f(x, y)$ dada por la siguiente expresión:

$$f(x, y) = \begin{cases} e^{-(x+y)} & \text{si } x, y > 0, \\ 0 & \text{otro caso.} \end{cases}$$

a) Grafique $f(x, y)$ y demuestre que efectivamente se trata de una función de densidad.

b) Calcule y grafique las densidades marginales $f_1(x)$ y $f_2(y)$. Verifique que ambas son efectivamente funciones de densidad.

c) Calcule y grafique las distribuciones marginales $F_1(x)$ y $F_2(y)$. Verifique que ambas son efectivamente funciones de distribución.

d) ¿Son X y Y independientes?

Covarianza

255. A partir de la definición de covarianza, demuestre que

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y).$$

256. Demuestre las siguientes dos propiedades de la covarianza. Esto comprueba que la covarianza es lineal en cada argumento.
   a) $\text{Cov}(c\,X, Y) = c\ \text{Cov}(X, Y)$, c constante.
   b) $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$.

257. Sean X y Y dos variables aleatorias discretas con función de probabilidad conjunta dada por la tabla que aparece abajo. Compruebe que la covarianza entre X y Y es $1/2$.

x\y	-1	1
-1	3/8	1/8
1	1/8	3/8

258. Sean X y Y dos variables aleatorias continuas con función de densidad conjunta dada por la expresión que aparece abajo. Compruebe que la covarianza entre X y Y es cero.

$$f(x, y) = \begin{cases} 4xy & \text{si } 0 \leq x, y \leq 1, \\ 0 & \text{otro caso.} \end{cases}$$

Coeficiente de correlación

259. Compruebe que el coeficiente de correlación entre las dos variables aleatorias discretas X y Y con la siguiente función de probabilidad conjunta es cero.

x\y	0	1
0	1/4	1/4
1	1/4	1/4

260. El siguiente ejemplo muestra dos variables aleatorias cuya covarianza es cero (y por lo tanto el coeficiente de correlación también lo es) y sin embargo las variables aleatorias no son independientes: sea X con distribución unif$\{-1, 0-1\}$ y sea $Y = X^2$. Demuestre que:
   a) $\text{Cov}(X, Y) = 0$.
   b) X y Y no son independientes.

261. Sean $X_1, X_2, \ldots, X_n$ variables aleatorias independientes con idéntica distribución Ber(p). Demuestre que
   a) $\text{Cov}(X_1, X_1 + \cdots + X_n) = p(1-p)$.
   b) $\rho(X_1, X_1 + \cdots + X_n) = 1/\sqrt{n}$.

Muestras aleatorias y estadísticas

262. Sea X_1, X_2, X_3 una muestra aleatoria de la distribución Ber(p). Encuentre la distribución, media y varianza de la estadística

$$X = X_1 + X_2 + X_3.$$

263. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución $\mathrm{N}(\mu, \sigma^2)$, en donde los parámetros μ y σ^2 son desconocidos. Diga si las siguientes variables aleatorias son estadísticas:

a) $T_1 = 4(X_1 + \cdots + X_n)$.
b) $T_2 = (X_1 + \cdots + X_n) - \mu$.
c) $T_3 = \sum_{i=1}^{n} (X_i - X_{i-1})^2$.
d) $T_4 = \sigma^2 + X_1 + X_2$.
e) $T_5 = X_n$.
f) $T_6 = \text{máx}\{X_1, \ldots, X_n\}$.

264. Sea $X_1, \ldots, X_n$ una muestra aleatoria de una población con distribución $\mathrm{N}(\mu, \sigma^2)$. Encuentre la distribución, media y varianza de la estadística

$$S = X_1 + \cdots + X_n.$$

265. Calcule e interprete el coeficiente de correlación para las calificaciones en las siguientes dos asignaturas de cinco estudiantes escogidos al azar.

Calif. en literatura	8.5	7.0	10.0	8.5	9.0
Calif. en matemáticas	8.0	8.0	9.0	8.5	8.5

Estimación puntual

266. Sea $X_1, \ldots, X_n$ una m.a. de una población con función de probabilidad $\mathrm{Ber}(p)$. Calcule la esperanza y la varianza del estimador puntual

$$\hat{p} = \bar{X}.$$

267. Sea $X_1, \ldots, X_n$ una m.a. de una población con función de densidad $\exp(\lambda)$. Calcule la esperanza y la varianza de los siguientes estimadores para λ.
a) $\hat{\lambda}_1 = X_1 + \cdots + X_n$.
b) $\hat{\lambda}_2 = \frac{1}{2}(X_1 + X_2)$.
c) $\hat{\lambda}_3 = \bar{X}$.

Método de momentos

268. Dada una muestra aleatoria de tamaño n de una población uniforme en el intervalo $[0, a]$, use el método de momentos para encontrar un estimador para el parámetro a.
269. Sea $X_1, \ldots, X_n$ una muestra aleatoria con función de densidad uniforme en el intervalo (a, b). Obtenga los estimadores para a y b por el método de momentos.
270. Se cuenta con 5 observaciones independientes de una variable aleatoria X con distribución uniforme en el intervalo $(a - b, a + b)$, en donde $a \in \mathbb{R}$ y $b > 0$,

$$20, 21.5, 19, 20.3, 21\,.$$

a) Use el método de momentos para estimar los valores de los parámetros a y b.
b) Habiendo estimado los parámetros a y b, calcule $P(X > 21)$.

271. El tiempo aleatorio en minutos para la ocurrencia de un cierto evento de interés se modela mediante una variable aleatoria X con distribución exponencial de parámetro desconocido $\lambda > 0$. Se efectúan 6 observaciones independientes de esta variable aleatoria obteniéndose los valores

$$15, 12, 25, 16, 20, 19\,.$$

a) Use el método de momentos para encontrar un estimador del parámetro λ.
b) Calcule el tiempo medio de espera para la ocurrencia del evento en estudio.

272. Dada una muestra aleatoria de tamaño n de una población Poisson con parámetro desconocido $\lambda > 0$, use el método de momentos para encontrar un estimador del parámetro λ.

273. La variación V de ciertas mediciones se modela mediante una variable aleatoria con distribución $\mathrm{N}(0, \sigma^2)$, en donde σ^2 es un parámetro desconocido.
a) Con base en la muestra $0.2, -0.5, -0.3, 0.4$ estime el valor de σ^2 por el método de momentos.
b) Calcule la probabilidad de que la variación exceda en valor absoluto el nivel 0.2.

274. Una variable aleatoria de interés X puede tomar los valores $-1, 0, 1$. Estos valores pueden corresponder a las categorías: malo, regular, bueno. Se postula que esta variable aleatoria tiene la siguiente función de probabilidad:

$$f(x) = \begin{cases} (\frac{\theta}{2})^{|x|}(1-\theta)^{1-|x|} & \text{si } x = -1, 0, 1, \\ 0 & \text{otro caso,} \end{cases}$$

en donde $\theta \in [0, 1]$ es un parámetro. Use el método de momentos para estimar θ a partir de la muestra particular: $-1, -1, 0, 1, 1, 0, 0$.

275. Sean $f_1(x)$ y $f_2(x)$ dos funciones de densidad conocidas con esperanzas distintas μ_1 y μ_2, respectivamente. Para cada $\theta \in [0, 1]$ defina la función de densidad

$$f(x) = \theta\, f_1(x) + (1-\theta)\, f_2(x).$$

Sean $x_1, x_2, \ldots, x_n$ observaciones independientes provenientes de la densidad $f(x)$. Estime el parámetro θ por el método de momentos.

Método de máxima verosimilitud

276. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de la distribución $\mathrm{Ber}(p)$. Demuestre que el estimador máximo verosímil para el parámetro desconocido p está dado por $\hat{p} = \bar{X}$.

277. Dada una muestra aleatoria de tamaño n de una población Poisson con parámetro $\lambda > 0$, use el método de máxima verosimilitud para encontrar un estimador del parámetro λ.

278. Sea $X_1, \ldots, X_n$ una m.a. de una población con función de probabilidad

$$f(x) = p(1-p)^{x-1} \quad \text{para} \quad x = 1, 2, \ldots$$

en donde $p \in (0, 1)$ es un parámetro. Use el método de máxima verosimilitud para encontrar un estimador para p.

279. Sea $X_1, \ldots, X_n$ una m.a. de una población con función de densidad

$$f(x) = \theta x^{\theta-1} \quad \text{para} \quad 0 < x < 1,$$

en donde $\theta \in (0, \infty)$ es un parámetro. Use el método de máxima verosimilitud para encontrar un estimador para θ.

280. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de la distribución uniforme en el intervalo $[a, b]$. Justifique que los estimadores máximo verosímiles para los parámetros a y b pueden ser

$$\begin{aligned} \hat{a} &= \text{mín}\,\{X_1, X_2, \ldots, X_n\}, \\ \hat{b} &= \text{máx}\,\{X_1, X_2, \ldots, X_n\}. \end{aligned}$$

281. Sea $X_1, X_2, \ldots, X_m$ una muestra aleatoria de tamaño m de la distribución bin(n, p). Justifique que los estimadores máximo verosímiles para los parámetros n y p pueden ser

$$\begin{aligned} \hat{n} &= \text{máx}\,\{X_1, X_2, \ldots, X_m\}, \\ \hat{p} &= \bar{X}/\hat{n}. \end{aligned}$$

282. Usando una muestra de tamaño n, encuentre el estimador máximo verosimil para el parámetro λ de la distribución Weibull(α, λ). Considere que el parámetro α es conocido.

$$f(x) = \begin{cases} e^{-(\lambda x)^\alpha} \lambda\alpha\, (\lambda x)^{\alpha-1} & \text{si } x > 0, \\ 0 & \text{otro caso.} \end{cases}$$

283. El tiempo de vida útil en años de un cierto dispositivo se modela como una variable aleatoria X con función de densidad

$$f(x) = \begin{cases} 3\,\dfrac{a^3}{x^4} & \text{si } x \geq a, \\ 0 & \text{otro caso.} \end{cases}$$

a) Encuentre el estimador máximo verosimil para el parámetro a, con base en la muestra de cuatro mediciones: 10.5, 9.3, 11, 11.7 .

b) Calcule el tiempo medio de vida útil de uno de estos dispositivos.

284. Usando una muestra de tamaño m, encuentre el estimador máximo verosimil para el parámetro λ de la distribución gama(n, λ). Considere que el parámetro n es conocido.

285. Considere la función de densidad continua

$$f(x) = \begin{cases} \dfrac{2x}{\theta^2} & \text{si } 0 < x < \theta, \\ 0 & \text{otro caso.} \end{cases}$$

a) Compruebe que $f(x)$ es efectivamente una función de densidad.

b) Sean $1.7, 1.5, 3, 2.1, 2.5$ observaciones independientes de esta función de densidad. Use el método de máxima verosimilitud para estimar el valor del parámetro θ con base en esta muestra.

c) Usando el valor estimado para θ, calcule el valor promedio de una v.a. X con esta función de densidad.

286. El tiempo aleatorio X que es necesario esperar para que ocurra un cierto evento de interés siempre es mayor a una constante θ y tiene la siguiente función de densidad

$$f(x) = \begin{cases} e^{-(x-\theta)} & \text{si } x \geq \theta, \\ 0 & \text{otro caso.} \end{cases}$$

a) Compruebe que $f(x)$ es efectivamente una función de densidad.

b) Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de esta distribución. Encuentre el estimador máximo verosimil para el parámetro desconocido θ.

c) Estime θ con base en la muestra $7, 5, 9, 6, 5, 8$.

d) Encuentre el valor promedio estimado de este tiempo aleatorio.

Insesgamiento

287. Sean $\hat{\theta}_1$ y $\hat{\theta}_2$ dos estimadores insesgados para un parámetro θ.

a) Sea a una constante. Demuestre que el siguiente estimador también es insesgado para θ:

$$\hat{\theta} = a\,\hat{\theta}_1 + (1-a)\,\hat{\theta}_2.$$

b) Encuentre condiciones sobre las constantes a y b para que el siguiente estimador sea insesgado:

$$\hat{\theta} = a\,\hat{\theta}_1 + b\,\hat{\theta}_2.$$

288. Sea $X_1, \ldots, X_n$ una m.a. de una población con función de densidad de probabilidad unif$(0, \theta)$. Determine si los siguientes estimadores son insesgados para θ.

a) $\hat{\theta}_1 = \bar{X}$.

b) $\hat{\theta}_2 = 2\bar{X}$.

c) $\hat{\theta}_3 = \dfrac{1}{2}(X_1 + X_n)$.

289. Sea $X_1, \ldots, X_n$ una m.a. de una distribución cualquiera con media μ y varianza finita σ^2. Demuestre que las siguientes estadísticas son estimadores insesgados para σ^2.

a) $\hat{\sigma}^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \mu)^2$, suponiendo μ conocida.

b) $S^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2$.

c) $\hat{\sigma}^2 = \dfrac{1}{2(n-1)} \sum_{i=1}^{n-1} (X_{i+1} - X_i)^2$.

290. Sea $X_1, \ldots, X_n$ una m.a. de una población con media μ y varianza σ^2 desconocidas. Recuerde que la varianza muestral S^2 se ha definido como sigue:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2.$$

a) Demuestre que S^2 es un estimador insesgado para σ^2.
b) Demuestre que S no es un estimador insesgado para σ. Modifique S para que sea insesgado.

291. Sea $X_1, \ldots, X_n$ una m.a. de una población con función de densidad $\mathrm{N}(0, \theta)$ y sea $\hat{\theta} = (X_1^2 + \cdots + X_n^2)/n$. Demuestre que:
a) $E(\hat{\theta}) = \theta$.
b) $\mathrm{Var}(\hat{\theta}) = 2\theta^2/n$.

292. Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de la función de densidad

$$f(x) = \theta\, f_1(x) + (1-\theta)\, f_2(x),$$

en donde $\theta \in [0, 1]$ y $f_1(x)$ y $f_2(x)$ son dos funciones de densidad con medias distintas μ_1 y μ_2, respectivamente. Demuestre que el siguiente estimador es insesgado para θ:

$$\hat{\theta} = \frac{\bar{X} - \mu_2}{\mu_1 - \mu_2}.$$

Estimación por intervalos

293. Se sabe que la vida en horas de un foco de 100 watts de cierta marca tiene una distribución aproximada normal con desviación estándar $\sigma = 30$ horas. Se tomó una muestra al azar de 50 focos y resultó que la vida media fue de 1550 horas. Construya un intervalo de confianza del 95 % para el verdadero promedio de vida útil de estos focos.

294. Se realizan 20 pruebas de resistencia de un cierto material obteniéndose los siguientes datos: 2225, 2300, 2217, 2190, 2295, 2285, 2195, 2255, 2232, 2252, 2272, 2231, 2223, 2211, 2219, 2231, 2218, 2262, 2257, 2261. Construya un intervalo de confianza del 98 % para la resistencia media de este material, suponiendo una distribución normal.

295. Suponga que se tiene una muestra aleatoria con función de densidad de probabilidad $\mathrm{Ber}(p)$. Se desea obtener un intervalo de confianza al 90 % para p de tal forma que la longitud del intervalo sea a lo sumo 0.05. ¿Qué tamaño de muestra debe tomarse?

296. En las pruebas de resistencia de cierto tipo de cables, se ha encontrado que 10 de ellos se han roto cuando se les ha aplicado la siguiente tensión en kilo

Newtons

$$100, 105, 98, 130, 125, 120, 110, 108, 110, 105.$$

Suponiendo una desviación estándar de 20kN, encuentre un intervalo de confianza al 95 % para la tensión media con la que este tipo de cables se rompe.

Pruebas de hipótesis

297. Considere el Ejemplo 2.52 de la página 128 sobre la prueba de hipótesis para determinar si una moneda es equilibrada. Determine el número de cruces que se deben obtener para rechazar o no rechazar la hipótesis nula.
298. Suponga que se tiene una sola observación X de una población con distribución uniforme en el intervalo $(0, \theta)$. Suponga que se desea llevar a cabo la prueba de hipótesis

$$H_0 : \theta = 1 \quad vs \quad H_1 : \theta > 1.$$

a) Calcule la probabilidad de cometer el error tipo I cuando la región de rechazo está dada por el intervalo $(9/10, \infty)$.
b) Para cada valor de $\theta_1 > 1$ fijo calcule la probabilidad de cometer el error tipo II, $\beta(\theta_1)$.

Pruebas para la media de una distribución normal con varianza conocida

299. Se cree que la estatura de cierta población humana sigue una distribución normal con media de 1.70 metros. Realice la prueba de hipótesis

$$H_0 : \mu = 1.70 \quad vs \quad H_1 : \mu \neq 1.70,$$

con el siguiente conjunto de datos: 1.65, 1.75, 1.63, 1.81, 1.74, 1.59, 1.73, 1.66, 1.65, 1.83, 1.77, 1.74, 1.64, 1.69, 1.72, 1.66, 1.55, 1.60, 1.62. Use un nivel de significancia del 10 % y suponga $\sigma = 0.10$.
300. Las mediciones del número de cigarros fumados al día por un grupo de diez fumadores es el siguiente: $5, 10, 3, 4, 5, 8, 20, 4, 1, 10$. Realice la prueba de hipótesis

$$H_0 : \mu = 10 \quad vs \quad H_1 : \mu < 10,$$

suponiendo que los datos provienen de una muestra tomada al azar de una población normal con $\sigma = 1.2$. Use un nivel de significancia del 5 %.
301. Considere cualquiera de las tres pruebas de hipótesis para la media μ de una población normal con σ^2 conocida. Demuestre que en cualquier caso la probabilidad de cometer el error tipo II tiende a cero cuando el tamaño de la muestra n tiende a infinito.

Pruebas para la media de una distribución normal con varianza desconocida

302. Una compañía que produce cierto tipo de bebidas está utilizando una nueva maquinaria en el llenado de sus envases de 600ml. La nueva máquina es más rápida que las anteriores pero se desea saber si debido a la rapidez no se pierde precisión en el llenado. Se toman al azar 100 envases llenados por la nueva máquina y se mide con cuidado el contenido. El promedio fue $\bar{x} = 601.5$ml. y la varianza muestral fue $s^2 = 5.2$. Con un nivel de significancia del 5 %, lleve a cabo la prueba de hipótesis

$$H_0 : \mu = 600\text{ml.} \quad vs \quad H_1 : \mu \neq 600\text{ml.}$$

303. Por la falta de suficiente alimento, se cree que la longitud de cierta especie de peces en un lago ha decrecido de su promedio habitual que es de 20cm. Con base en la muestra

$$20.3, 19, 18.5, 20, 19.5, 17, 19.3, 18.4, 19.5, 21,$$

y con un nivel de significancia del 10 % lleve a cabo la prueba de hipótesis

$$H_0 : \mu = 20\text{cm} \quad vs \quad H_1 : \mu < 20\text{cm.}$$

304. El número de preguntas promedio contestadas correctamente en un examen de 100 preguntas es de 65. Una escuela ofrece un curso de preparación para presentar dicho examen y garantiza un resultado promedio mayor para aquellas personas que toman el curso. Con base en la muestra

$$62, 80, 55, 20, 97, 70, 68, 64, 71, 55, 57, 60, 88,$$

de resultados de personas que han presentado el examen después de haber tomado el curso, y para un nivel de significancia del 10 % lleve a cabo la prueba de hipótesis

$$H_0 : \mu = 65 \quad vs \quad H_1 : \mu > 65.$$

Apéndice B

Solución a los ejercicios

Esta sección contiene algunas sugerencias para encontrar la solución a los ejercicios planteados. Para algunos ejercicios es necesario ser más explícito al dar una solución o al justificar una respuesta, considere por tanto que este material contiene simplemente ideas para generar una solución completa y bien escrita. La mayoría de las gráficas han sido omitidas. Recuerde además que los métodos empleados o sugeridos para llegar a una solución no son necesariamente únicos.

Espacio muestral y eventos

1. *a*) $\Omega = \{AAA, AAS, ASA, ASS, SAA, SAS, SSA, SSS\}$.
   b) $\Omega = \{AAA, AAS, ASS, SSS\}$.
   c) $\Omega = [0, 1]$.
   d) $\Omega = \{0, 1, 2, \ldots\}$.
   e) Considere que las bolas tienen las etiquetas 1 y 2. Considere el arreglo

   $$(C_1, C_2, C_3, C_4)$$

   en donde cada coordenada C_i corresponde al conjunto de bolas que han sido depositadas en la celda i. Entonces Ω es el conjunto

   $$\{(\{1,2\}, \emptyset, \emptyset, \emptyset), (\emptyset, \{1,2\}, \emptyset, \emptyset), \ldots, (\{1\}, \{2\}, \emptyset, \emptyset), \ldots\}.$$

   El número total de elementos en Ω es 16. ¿Puede usted escribir todos los arreglos?
   f) $\Omega = \{(2,0,0,0), (0,2,0,0), (0,0,2,0), (0,0,0,2), (1,1,0,0), (1,0,1,0), (1,0,0,1), (0,1,1,0), (0,1,0,1), (0,0,1,1)\}$.
2. $\Omega = \{(0,0), (1,0), (0,1), (1,1), \ldots\}$.
3. Suponiendo que las computadoras se encuentran numeradas,

   $$\Omega = \{(x_1, \ldots, x_{100}) : x_i \in \{0, 1\} \text{ para } i = 1, \ldots, 100\}.$$

4. Si nos interesa observar el número de lanzamientos hasta obtener un 6 se puede tomar $\Omega = \{1, 2, \ldots\}$. Si nos interesa registrar con detalle los resultados, $\Omega = \{6, N6, NN6, NNN6, \ldots\}$, en donde cada N puede ser cualquiera de los números $1, 2, 3, 4$ ó 5.
5. a) Número de hijos de una persona escogida al azar. b) Tiempo que una persona tiene que esperar para la llegada de un autobús. c) Este es el espacio muestral más frecuente. El número 1 puede representar ganar en un juego de lotería mientras que el número 0 corresponde a perder en el juego. Este es uno de los muchos ejemplos que pueden construirse para este espacio.

Operaciones con conjuntos

6. Sea x en B^c. Entonces $x \notin B$. Por lo tanto $x \notin A$, es decir, $x \in A^c$.
7. Sea x en A. Como $A = A \cap B$, x también pertenece a B. Por lo tanto todo x elemento de A es también un elemento de B, es decir, $A \subseteq B$.
8. Sea x en $B - A$. Si tal elemento existe, entonces $x \notin A$, es decir, $x \in A^c$.
9. a) Sea $x \in A \cap (B \cup C)$. Entonces $x \in A$ y $x \in B \cup C$. La última afirmación puede descomponerse en: $x \in B$ ó $x \in C$, esto incluye el caso de pertenencia a ambos conjuntos. Si es el primer caso, es decir, si $x \in B$, entonces se obtiene que $x \in A \cap B$, y por lo tanto x es un elemento del lado derecho de la igualdad. La misma conclusión se obtiene si $x \in C$. Esto demuestra la contención $A \cap (B \cup C) \subseteq (A \cap B) \cup (A \cap C)$. De manera análoga se prueba la contención contraria. b) Análogo al inciso anterior.
10. a) Demostraremos la contención $(A \cup B)^c \subseteq A^c \cap B^c$. Sea $x \in (A \cup B)^c$. Entonces $x \notin (A \cup B)$, esto es, $x \notin A$ y $x \notin B$. Por lo tanto $x \in A^c$ y $x \in B^c$. Es decir, $x \in A^c \cap B^c$. De manera análoga se prueba la contención contraria y de esa forma se demuestra la igualdad. b) Pruebe ambas contenciones como en el inciso anterior.
11. Los enunciados son:
    a) $(A \cup B \cup C)^c = A^c \cap B^c \cap C^c$.
    b) $b)(A \cap B \cap C)^c = A^c \cup B^c \cup C^c$.
    Para el primer inciso se tiene que $(A \cup B \cup C)^c = ((A \cup B) \cup C)^c = (A \cup B)^c \cap C^c = A^c \cap B^c \cap C^c$. De manera análoga se demuestra al segundo inciso.
12. *a)* $A = A \cap \Omega = A \cap (B \cup B^c) = (A \cap B) \cup (A \cap B^c)$.
    b) $A \cap (B \cup A^c) = (A \cap B) \cup (A \cap A^c) = (A \cap B) \cup \emptyset = A \cap B$.
    c) $A \cup (B \cap A^c) = (A \cup B) \cap (A \cup A^c) = (A \cup B) \cap \Omega = A \cup B$.
13. *a)* $A - (A \cap B) = A \cap (A \cap B)^c = A \cap (A^c \cup B^c) = (A \cap A^c) \cup (A \cap B^c) = \emptyset \cup (A \cap B^c) = A \cap B^c = A - B$.
    b) $(A \cup B) - B = (A \cup B) \cap B^c = (A \cap B^c) \cup (B \cap B^c) = (A \cap B^c) \cup \emptyset = A \cap B^c = A - B$.
14. *a)* $(A \cap B) - (A \cap C) = (A \cap B) \cap (A \cap C)^c = (A \cap B) \cap (A^c \cup C^c) = (A \cap B \cap A^c) \cup (A \cap B \cap C^c) = \emptyset \cup (A \cap B \cap C^c) = A \cap (B \cap C^c) = A \cap (B - C)$.
    b) $(A - C) \cap (B - C) = (A \cap C^c) \cap (B \cap C^c) = A \cap (B \cap C^c) = (A \cap B) - C$.
15. Empezando con la segunda identidad, $A \triangle B = (A \cup B) - (A \cap B) = (A \cup B) \cap (A \cap B)^c = (A \cup B) \cap (A^c \cup B^c) = [(A \cup B) \cap A^c] \cup [(A \cup B) \cap B^c] = [B \cap A^c] \cup [A \cap B^c] = (B - A) \cup (A - B)$.
16. Compruebe que ambos lados de la igualdad del inciso (a) corresponden al evento que se muestra en la Figura B.1. El resto de los incisos se obtiene fácilmente de la definición.
17. El conjunto A es el intervalo $[-1, 1]$, B es $(0, \infty)$. Entonces $A \cup B = [-1, \infty)$, $A \cap B = (0, 1]$, $A - B = [-1, 0]$ y $B - A = (1, \infty)$.
18. Gráficamente los conjuntos A y B son como se muestran en la Figura B.2. Se tiene entonces que $A = [-2, 6]$, $B = (-\infty, -1) \cup (3, \infty)$, $A^c = (-\infty, -2) \cup (6, \infty)$, $B^c = [-1, 3]$, $A \cap B = [-2, -1) \cup (3, 6]$, $A \cup B = \mathbb{R}$, $A - B = [-1, 3]$, $B - A = (-\infty, -2) \cup (6, \infty)$, $A \triangle B = (-\infty, -2) \cup [-1, 3] \cup (6, \infty)$.
19. Gráficamente los conjuntos A y B son los que se muestran en la Figura B.3. Por lo tanto $A = [-3, 1]$, $B = (-\infty, -\sqrt{2}) \cup (\sqrt{2}, \infty)$, $A^c = (-\infty, -3) \cup (1, \infty)$, $B^c =$

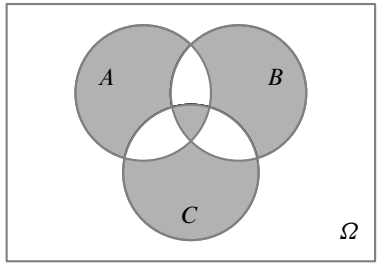


FIGURA B.1. $A \triangle B \triangle C$.

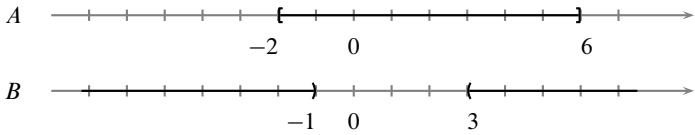


FIGURA B.2. Intervalos.

$[-\sqrt{2}, \sqrt{2}]$, $A \cap B = [-3, -\sqrt{2})$, $A \cup B = (-\infty, 1) \cup (\sqrt{2}, \infty)$, $A - B = [-\sqrt{2}, 1]$, $B - A = (-\infty, -3) \cup (\sqrt{2}, \infty)$, $A \triangle B = (-\infty, -3) \cup [-\sqrt{2}, 1] \cup (\sqrt{2}, \infty)$.

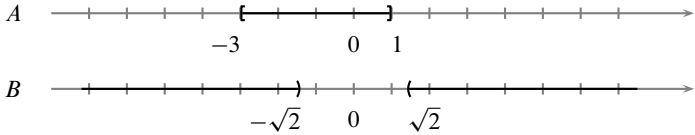
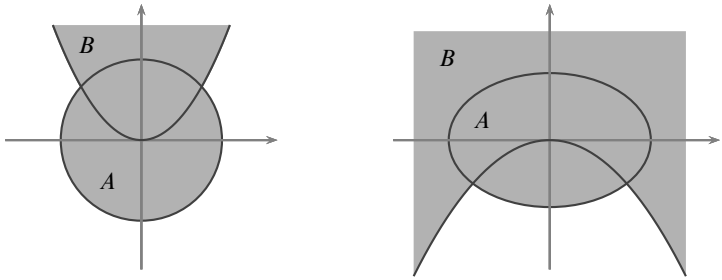


FIGURA B.3. Intervalos.

20. Los conjuntos A y B son los que se muestra en la Figura B.4 (a). Allí pueden identificarse gráficamente las operaciones que se solicitan.
21. Los conjuntos A y B son los que se muestran en la Figura B.4 (b). Allí pueden identificarse gráficamente las operaciones que se solicitan.
22. El área del conjunto $A \cap B$ es 1.
23. El máximo es 50 % y el mínimo es 0 %.
24. El máximo es 70 % y el mínimo es 40 %.
25. a) $A \cap B \cap C$. b) $[A \cap B \cap C^c] \cup [A \cap B^c \cap C] \cup [A^c \cap B \cap C]$.
    c) $[A \cap B^c \cap C^c] \cup [A^c \cap B \cap C^c] \cup [A^c \cap B^c \cap C]$.
    d) $A^c \cap B^c \cap C^c$. e) $A \cap B^c \cap C$.

Conjuntos ajenos

26. $A \cap (B - A) = A \cap (B \cap A^c) = \emptyset$.



(a) (b)

FIGURA B.4. Regiones.

27. Sea x en A, entonces x no puede pertenecer a B pues $A \cap B = \emptyset$. Por lo tanto x pertenece a B^c. Esto demuestra que $A \subseteq B^c$. La otra contención es análoga.
28. Las raíces de la ecuación cuadrática son $x = -1, 2$. Por lo tanto los conjuntos A y B no son ajenos.
29. Escriba la expresión matemática para la intersección de cualesquiera dos de estos conjuntos. En esta expresión aparecerá el término $A \cap A^c$ o bien $B \cap B^c$, el cual es vacío y eso hace que la intersección total sea el conjunto vacío.
30. No son ajenos, tienen un único punto en común que es $(0, 1)$.
31. $A_1 \cap B \subseteq A \cap B = \emptyset$.
32. $n(n-1)/2$.

Conjunto potencia

33. $2^{\emptyset} = \{\emptyset\}$.
34. $2^{\Omega} = \{\emptyset, \{a\}\}$.
35. Sea $A_1 \in 2^A$, entonces $A_1 \subseteq A \subseteq B$. Por lo tanto $A_1 \subseteq B$ y en consecuencia $A_1 \in 2^B$.
36. No, pues 2^A no es un número.

Producto Cartesiano

37. $A \times B \times C = \{(a_1, b_1, c_1), (a_1, b_1, c_2), (a_1, b_2, c_1), (a_1, b_2, c_2),(a_1, b_3, c_1), (a_1, b_3, c_2),(a_2, b_1, c_1),(a_2, b_1, c_2), (a_2, b_2, c_1),(a_2, b_2, c_2), (a_2, b_3, c_1), (a_2, b_3, c_2)\}$.
38. $\Omega \times \Omega = \{(a, a),(a, b),(a, c),(b, a),(b, b),(b, c),(c, a),(c, b),(c, c)\}$.
39. 2^n.

Probabilidad

40. Claramente $P(A) = \#A/\#\Omega \geq 0$, y $P(\Omega) = \#\Omega/\#\Omega = 1$. Además, cuando A y B son ajenos, $\#(A \cup B) = \#A + \#B$, y de allí se sigue la tercera propiedad.

41. Para cualquier valor natural de n, se cumple que $n(A)/n \geq 0$, y $n(\Omega)/n = n/n = 1$. Cuando A y B son ajenos, $n(A \cup B) = n(A) + n(B)$. Ahora tome límite cuando n tiende a infinito y obtenga el resultado requerido.
42. *a*) $P(\Omega) = \alpha P_1(\Omega) + (1-\alpha)P_2(\Omega) = \alpha + (1-\alpha) = 1$.
    b) $P(A) = \alpha P_1(A) + (1-\alpha)P_2(A) \geq 0$.
    c) Cuando A y B son disjuntos, $P(A \cup B) = \alpha P_1(A \cup B) + (1-\alpha)P_2(A \cup B) = \alpha(P_1(A) + P_1(B)) + (1-\alpha)(P_2(A) + P_2(B)) = [\alpha P_1(A) + (1-\alpha)P_2(A)] + [\alpha P_1(B) + (1-\alpha)P_2(B)] = P(A) + P(B)$.
43. $P(\emptyset) = P(\emptyset \cup \emptyset) = P(\emptyset) + P(\emptyset)$. Ahora cancele uno de los términos.
44. $B = (B \cap A) \cup (B \cap A^c)$, siendo esta unión ajena. Aplicando probabilidad, $P(B) = P(B \cap A) + P(B \cap A^c)$. El segundo sumando es $P(B - A)$.
45. La primera desigualdad es evidente. Como $A \cap B \subseteq A$, se tiene que $P(A \cap B) \leq P(A)$. Análogamente como $A \subseteq A \cup B$, $P(A) \leq P(A \cup B)$. La siguiente desigualdad es consecuencia de la fórmula $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Finalmente la última desigualdad se obtiene de $P(A) \leq P(\Omega) = 1$.
46. El resultado se obtiene de $1 \geq P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
47. a) $P(A \cup B) = 0.5$. b) $P(A^c) = 0.7$. c) $P(A^c \cap B) = 0.2$. d) $P(A \cap B^c) = 0.3$. e) $P(A \triangle B) = 0.5$.
48. a) Verdadero, pues $A \cap B \subseteq A$. b) Falso, considere por ejemplo el experimento de lanzar una moneda equilibrada. Ambos resultados tienen la misma probabilidad pero evidentemente no son iguales. c) Falso, en el experimento de lanzar un dado equilibrado, considere por ejemplo los eventos $A = \{1\}$ y $B = \{2, 3\}$. Claramente $P(A) \leq P(B)$ pero A no está contenido en B.
49. *a*) Verdadero, pues $A \subseteq A \cup B$.
    b) Falso, A y B pueden ser ajenos.
    c) Verdadero, $P(A \cup B) = P(A) + P(B) - P(A \cap B) > 1/2 + 1/2 - P(A \cap B) = 1 - P(A \cap B) \geq 0$.
    d) Falso, A puede ser Ω.
50. *a*) Falso. En el experimento de lanzar una moneda equilibrada, sea A el resultado de obtener una de las caras y B obtener la otra cara. Entonces $P(B - A) = P(B) = 1/2$, mientras que $P(B) - P(A) = 1/2 - 1/2 = 0$. El resultado es cierto bajo la condición $A \subseteq B$.
    b) Falso, excepto cuando A y B son independientes. Considerando el ejemplo del inciso anterior, $P(A \cap B) = 0$, mientras que $P(A)P(B) = 1/2 \cdot 1/2 = 1/4$.
    c) Verdadero. $P(A^c) = 1 - P(A) < 1 - 1/2 = 1/2$.
    d) Falso, A puede ser Ω.
    e) Cierto pues $A \subseteq B^c$.
51. $P(A) = 2/3$.
52. Siguiendo la definición de diferencia simétrica,
    $P(A \triangle B) = P((A \cup B) - (A \cap B)) = P(A \cup B) - P(A \cap B) = P(A) + P(B) - P(A \cap B) - P(A \cap B) = P(A) + P(B) - 2P(A \cap B)$.
53. Por la ley de De Morgan,
    $P(A^c \cap B^c) = P((A \cup B)^c) = 1 - P(A \cup B) = 1 - (P(A) + P(B) - P(A \cap B))$.

Análisis combinatorio

54. $6! = 720$.
55. *a*) $3^{10} = 59,049$. *b*) $\binom{12}{10} = 66$.
56. 16.
57. 380.
58. El total de casos posibles es $5! = 120$. Los casos favorables son dos: 12345 y 54321. Por lo tanto la probabilidad buscada es $2/120 = 1/60$.
59. $\binom{n}{2} - n = n(n-1)/2$.
60. $15 \cdot 14 \cdot 13 = 2730$.
61. Un entero positivo de a lo sumo cinco dígitos que es divisible por 2 es de la forma $xxxxy$ en donde x puede ser cualquiera de $0, 1, \ldots, 9$ y y puede ser $0, 2, 4, 6, 8$. Por lo tanto el total es $10 \cdot 10 \cdot 10 \cdot 10 \cdot 5 = 50000$, excepto que debemos omitir el entero 00000. Si empiezan con el número 1, entonces son de la forma $1xxxy$, y entonces la respuesta es $10 \cdot 10 \cdot 10 \cdot 5 = 5000$ y esta vez no hay necesidad de omitir ninguno de ellos.
62. $\binom{5}{3} = 10$.
63. $\binom{10}{2} = 45$.
64. El total de abrazos es $\binom{4}{2} = 6$. El total de saludos es $\binom{8}{2} - 4 = 28 - 4 = 24$. Por lo tanto el total de besos es $24 - 6 = 18$.
65. De $56!$ formas distintas.
66. a) $\binom{49}{6} = 13,983,816$ b) 6.
67. 70.
68. a) n^k b) n c) k^n d) $n!/(n-k)!$ e) $n!$ f) $n!$
69. En cada caso simplifique el lado derecho.
70. Considerando primero la situación en donde no hay empates, después cuando dos candidatos empatan, cuando tres candidatos empatan, cuando cuatro empatan y finalmente cuando todos empatan, la respuesta es $5! + 4! + 3! + 2! + 1! = 120 + 24 + 6 + 2 + 1 = 153$.
71. $\binom{50}{k}\binom{200}{20-k} / \binom{250}{20}$.
72. Considerando inicial y artificialmente que los siete dígitos son todos distintos, la respuesta preliminar es $7! = 5040$. Como los cuatro dígitos 1 son idénticos, cualquier permutación entre ellos produce el mismo número y por lo tanto debemos dividir el resultado preliminar por $4!$ Análogamente con los tres dígitos 0. La respuesta final es $7!/4!\,3! = 5040/(24 \cdot 6) = 35$.
73. Razonando como en el ejercicio anterior, la respuesta es $4!/(2!\,1!\,1!) = 6$.
74. Suponiendo que cada año tiene 365 días, la respuesta es $1 - 365 \cdot 364 \cdots (365 - n + 1)/365^n$.

75. El total es 4536. No es cierto que la mitad de ellos sean pares y la otra mitad impares. Existen 2296 pares y 2240 impares.
76. a) $\binom{10}{3}+\binom{15}{3}$. b) $\binom{25}{6}$. c) $\binom{15}{6}+\binom{10}{6}$.
77. Para $n=1$ se verifica directamente que $\#(2^{\Omega})=2=2^{\#\Omega}$. Suponga que la fórmula se cumple para un conjunto dado Ω_n de cardinalidad $n\geq 1$ y sea Ω_{n+1} el conjunto de $n+1$ elementos dados por los elementos de Ω_n y un elemento adicional x. Cada subconjunto de Ω_n genera dos subconjuntos de Ω_{n+1}: en un caso añadiendo el elemento adicional x y en el otro no añadiéndolo. Así, tenemos que $\#(2^{\Omega_{n+1}})=2\,\#(2^{\Omega_n})=2\,2^{\#\Omega_n}=2^{1+\#\Omega_n}=2^{\#\Omega_{n+1}}$.
78. Suponga $R(n)$ denota el número máximo de regiones con n rectas. Cuando no hay ninguna recta se tiene el plano completo y por lo tanto $R(0)=1$. Si se colocan una recta después de otra, entonces la n-ésima recta atraviesa siempre un máximo de n regiones. Por lo tanto se cumple la relación $R(n)=R(n-1)+n$. Resolviendo iterativamente se encuentra que $R(n)=R(0)+(1+2+\cdots+n)=1+n(n+1)/2$.
79. Siendo 5 y 7 números primos, los únicos divisores son aquellos números de la forma 5^x7^y para $x=0,1,2,3$ y $y=0,1,2,3,4$. Ello da un total de 20 divisores.
80. a) $\binom{20}{4}/\binom{40}{4}$. b) $\binom{20}{1}\binom{20}{3}/\binom{40}{4}$. c) $\binom{20}{2}\binom{20}{2}/\binom{40}{4}$.
    d) $\binom{20}{3}\binom{20}{1}/\binom{40}{4}$. e) $\binom{20}{4}/\binom{40}{4}$.
81. Se pueden obtener un total de 2^n subconjuntos distintos de un conjunto de cardinalidad n. Los 16 subconjuntos de $\{a,b,c,d\}$ son: $\emptyset$, $\{a\}$, $\{b\}$, $\{c\}$, $\{d\}$, $\{a,b\}$, $\{a,c\}$, $\{a,d\}$, $\{b,c\}$, $\{b,d\}$, $\{c,d\}$, $\{a,b,c\}$, $\{b,c,d\}$, $\{a,c,d\}$, $\{a,b,d\}$, $\{a,b,c,d\}$.
82. 2^n. El conjunto E corresponde al conjunto de los n ejes coordenados, o uniones de éstos.
83. a) 20. b) 400.
84. Compruebe que la fórmula se cumple para $n=1$. Suponga ahora que la fórmula es válida para cualquier $n\geq 1$ fijo. Con ayuda de esta hipótesis demuestre que la fórmula también se cumple para $n+1$ escribiendo y desarrollando $(a+b)^{n+1}=(a+b)^n(a+b)$.
85. 2^6.
86. $n!\,(n+1)!$
87. $(n!)^2$.
88. Denote por la letra a un paso hacia arriba y por d un paso a la derecha. Para llegar de A a B se necesitan 5 pasos hacia arriba y 7 pasos a la derecha. Así, la respuesta es el total de arreglos lineales en que se pueden acomodar las 12 letras $aaaaaddddddd$. La respuesta es entonces $12!/(5!\,7!)=792$.
89. $7!/(3!\,2!)=420$.
90. $11!/(5!\,4!\,2!)=6930$.
91. Considere que se tienen k casillas numeradas $1,2,\ldots,k$ en donde se desean colocar n bolas sin ninguna restricción en el número de bolas que caben en cada casilla. De esta forma, el número de bolas en la casilla 1 es el valor de x_1, lo mismo para las casillas

$2, 3, \ldots, k$. Las respuestas son entonces
a) $\binom{n+k-1}{n}$. b) $\sum_{m=0}^{n} \binom{m+k-1}{m}$. c) Infinito.

92. Considere que los números $1, 2, \ldots, n$ representan casillas que se encuentran ordenadas de manera natural. El problema puede entonces traducirse en colocar k bolas en este arreglo de casillas en donde cada casilla admite a lo sumo una bola. La respuesta es entonces $\binom{n}{k}$.
93. $(a+b+c)^3 = a^3+b^3+c^3+3ab^2+3ac^2+3a^2b+3a^2c+3b^2c+3bc^2+6abc$.
94. Substituya $a_1 = \cdots = a_m = 1$ en la fórmula del teorema multinomial.

Probabilidad condicional e independencia

95. *a*) Como $B = (A^c \cap B) \cup (A \cap B)$, se tiene que $P(A^c \cap B) = P(B) - P(A \cap B) = P(B) - P(A)P(B) = P(B)(1 - P(A)) = P(B)P(A^c)$.
b) Análogo al primer inciso.
c) $P(A^c \cap B^c) = 1 - P(A \cup B) = 1 - P(A) - P(B) + P(A \cap B) = 1 - P(A) - P(B) + P(A)P(B) = (1 - P(A))(1 - P(B)) = P(A^c)P(B^c)$.

96. *a*) $P(A \cap B) = 0.05$.
b) $P(A^c \cap B) = 0.45$.
c) $P(A \cap B^c) = 0.05$.
d) $P(A^c \cap B^c) = 0.45$.
e) $P(A \cup B) = 0.55$.
f) $P(A^c \cup B) = 0.95$.
g) $P(A \cup B^c) = 0.55$.
h) $P(A^c \cup B^c) = 0.95$.

97. a) $P(B) = 1/10$. b) $P(B) = 1/5$. c) $P(B) = 1/6$.
98. *a*) $P(A|B) \geq 0$.
b) $P(\Omega|B) = P(B)/P(B) = 1$.
c) Sean A_1 y A_2 ajenos. Entonces $A_1 \cap B$ y $A_2 \cap B$ también son ajenos. Por lo tanto $P(A_1 \cup A_2|B) = P((A_1 \cup A_2) \cap B)/P(B) = P((A_1 \cap B) \cup (A_2 \cap B))/P(B) = P(A_1 \cap B)/P(B) + P(A_2 \cap B)/P(B) = P(A_1|B) + P(A_2|B)$.
99. $P(A \cap B \cap C) = P(C|A \cap B)P(A|B)P(B) = ppp = p^3$.
100. *a*) Falso. Con $A = \emptyset$ se obtiene $P(A|B) = 0$.
b) b) Falso. Tómese $A = \Omega$ y B tal que $0 < P(B) < 1$. Entonces $P(A|B) = 1$ mientras que $P(B|A) = P(B) < 1$.
c) Falso. Tómese $A = B$ con $0 < P(A) < 1$. Entonces $P(A|B) = 1$ mientras que $P(A) < 1$.
101. *a*) Verdadero y ello es consecuencia del hecho de que $P(\cdot \mid B)$ es una medida de probabilidad. Alternativamente, $P(A \mid B) + P(A^c \mid B) = P(A \cap B)/P(B) + P(A^c \cap B)/P(B) = P((A \cap B) \cup (A^c \cap B))/P(B) = P((A \cup A^c) \cap B)/P(B) = P(B)/P(B) = 1$.
b) Falso. Tómese $A = \Omega$. El lado izquierdo es 2 mientras que el lado izquierdo es 1.
c) Cierto. $P(A \mid A \cap B) = P(A \cap B)/P(A \cap B) = 1$. El cálculo es idéntico para $P(B \mid A \cap B)$.
102. Todos los incisos son ciertos. Solo es cuestión de escribir la definición de cada probabilidad condicional.
103. *a*) Falso. Tómese por ejemplo $A = B_1 \cup B_2$. Entonces el lado izquierdo es uno mientras que el lado derecho es dos.

b) Cierto. $P(A_1 \cup A_2 \mid B) = P((A_1 \cup A_2) \cap B)/P(B) = P(A_1 \cap B)/P(B) + P(A_2 \cap B)/P(B) = P(A_1 \mid B) + P(A_2 \mid B)$.

c) Falso. Tómese $B_1 = B_2$. El lado derecho es el cuadrado del lado izquierdo.

d) Falso. Tómese $A_1 = A_2$. El lado derecho es el cuadrado del lado izquierdo.

104. *a*) $P(\emptyset \cap \emptyset) = P(\emptyset) = 0 = P(\emptyset)P(\emptyset)$.

b) $P(\Omega \cap \Omega) = P(\Omega) = 1 = P(\Omega)P(\Omega)$.

105. *a*) Falso. Tómese $B = \Omega$. Entonces A y B son independientes pero no son ajenos.

b) Falso. Considere el experimento de lanzar una moneda honesta, y sean A y B los eventos de que caiga una cara y la otra respectivamente. Entonces claramente A y B son ajenos pero no son independientes.

106. *a*) $P(A \cap \emptyset) = P(\emptyset) = 0 = P(A)P(\emptyset)$.

b) $P(A \cap \Omega) = P(A) = P(A)P(\Omega)$.

107. $P(A \cup B) = 1 - P(A^c \cap B^c) = 1 - P(A^c)P(B^c)$ pues A^c y B^c también son independientes. Alternativamente desarrolle $P(A \cup B)$ y factorice adecuadamente.

108. *a*) Cierto pues la definición es simétrica respecto de los eventos A y B.

b) Falso en general. Tómese A tal que $0 < P(A) < 1$. Entonces $P(A \cap A) = P(A)$ mientras que $P(A)P(A) < P(A)$. Cuando $P(A)$ es cero o uno la propiedad es válida.

c) Falso. Tómese $\Omega = \{1, 2, 3, 4\}$ equiprobable con $A = \{1, 2\}$, $B = \{2, 3\}$ y $C = \{3, 4\}$. Entonces se cumple $P(A \cap B) = 1/4 = P(A)P(B)$ y $P(B \cap C) = 1/4 = P(B)P(C)$, pero $P(A \cap C) = 0$ distinto a $P(A)P(C) = 1/4$.

109. La probabilidad de que no ocurra ninguno de los eventos es $P(A^c \cap B^c)$, que por la independencia es igual a $P(A^c)P(B^c) = (1 - p_1)(1 - p_2)$.

110. Como A y B son independientes y ajenos, $0 = P(A \cap B) = P(A)P(B)$. Para que este producto se anule forzosamente alguno de los factores tiene que ser cero.

111. a) $\emptyset = (0, 0), (0, 1/2)$. b) $(0, 1/2), (1/4, 3/4)$. c) $(0, 1/2), (1/2, 1)$. d) $(0, 1/2), (0, 1/2)$.

112. El total de subconjuntos de cardinal $1, 2, \ldots, n$ es, respectivamente, $\binom{n}{1}, \binom{n}{2}, \ldots, \binom{n}{n}$. De modo que la respuesta es $\binom{n}{1} + \cdots + \binom{n}{n} = 2^n - n - 1$.

113. Se puede usar el método de inducción sobre el valor de n. Alternativamente escriba la definición de probabilidad condicional para cada factor del lado derecho y simplifique.

114. $P(R_1 \cap R_2 \cap R_3) = P(R_1)P(R_2 \mid R_1)P(R_3 \mid R_2 \cap R_1)$. Por lo tanto, $P(R_1 \cap R_2 \cap R_3) = \frac{r}{r+c} \times \frac{r+c}{r+b+c} \times \frac{r+2c}{r+b+2c}$.

Teorema de probabilidad total

115. Se puede usar el método de inducción sobre n. Claramente $P(R_1) = r/(r + b)$. Suponga entonces válida la igualdad $P(R_n) = r/(r + b)$. Se condiciona la probabilidad del evento R_{n+1} dado el resultado de la primera extracción, y en tal situación puede considerarse que la configuración inicial de la urna cambia, de modo que el evento R_{n+1} se refiere ahora a la n-ésima extracción y allí es donde se usa la hipótesis de inducción. Entonces por el teorema de probabilidad total,

$$\begin{aligned} P(R_{n+1}) &= P(R_{n+1}|R_1)P(R_1) + P(R_{n+1}|R_1^c)P(R_1^c) \\ &= \frac{r+c}{r+b+c} \times \frac{r}{r+b} + \frac{r}{r+b+c} \times \frac{b}{r+c} \\ &= \frac{r}{r+b}. \end{aligned}$$

116. Sean N_1, N_2 y N_3 los eventos correspondientes a escoger los números 1, 2 y 3, respectivamente. Sea A el evento "obtener 5 al lanzar el dado". Entonces $P(A) = P(A \mid N_1)P(N_1) + P(A \mid N_2)P(N_2) + P(A \mid N_3)P(N_3)$, lo que resulta en el valor $11/108$.
117. $P(Y = 1) = P(Y = 1 \mid X = 1)P(X = 1) = (1/6)91/6)$,
$P(Y = 2) = P(Y = 2 \mid X = 1)P(X = 1) + P(Y = 2 \mid X = 2)P(X = 2) = (1/6 + 1/5)(1/6)$,
$P(Y = 3) = (1/6 + 1/5 + 1/4)(1/6)$,
$P(Y = 4) = (1/6 + 1/5 + 1/4 + 1/3)(1/6)$,
$P(Y = 5) = (1/6 + 1/5 + 1/4 + 1/3 + 1/2)(1/6)$,
$P(Y = 5) = (1/6 + 1/5 + 1/4 + 1/3 + 1/2 + 1)(1/6)$.

Teorema de Bayes

118. $1/4$.
119. Sea A el evento cuando todas las bolas son del mismo color, y N cuando en particular son todas negras. Entonces

$$P(N \mid A) = \frac{P(A \mid N)P(N)}{P(A)} = \frac{\binom{n}{k}}{\binom{n}{k} + \binom{b}{k}}.$$

120. $P(X = 2 \mid Y = 6) = \dfrac{P(Y = 6 \mid X = 2)P(X = 2)}{P(Y = 6)}$
$= \dfrac{(5/6)(1/6)}{(1/6 + 1/5 + 1/4 + 1/3 + 1/2 + 1)(1/6)} = 0.340\,.$
121. Sea R_0 el evento "se recibe un cero", y E_0 el evento "se envía un cero". Análogamente se definen los eventos R_1 y E_1. Entonces
   a) $P(R_0) = P(R_0|E_0)P(E_0) + P(R_0|E_1)P(E_1) = 0.8 \cdot 0.4 + 0.1 \cdot 0.6 = 0.38$.
   b) $P(R_1) = P(R_1|E_0)P(E_0) + P(R_1|E_1)P(E_1) = 0.2 \cdot 0.4 + 0.9 \cdot 0.6 = 0.62$.
   c) $P(E_0|R_0) = P(R_0|E_0)P(E_0)/P(R_0) = 0.8 \cdot 0.4/0.38 = 0.84$.
   d) $P(E_1|R_1) = P(R_1|E_1)P(E_1)/P(R_1) = 0.9 \cdot 0.6/0.62 = 0.87$.
122. *a)* $P(E) = P(E \mid B)P(B) + P(E \mid C)P(C) + P(E \mid D)P(D) = 0.361\,.$
   b) $P(F) = P(F \mid B)P(B) + P(F \mid C)P(C) + P(F \mid D)P(D) = 0.639\,.$
   c) $P(B \mid E) = P(E \mid B)P(B)/P(E) = 0.461\,.$
   d) $P(D \mid F) = P(F \mid D)P(D)/P(F) = 0.347\,.$
123. Los eventos de interés son $S =$ El estudiante sabe la respuesta, y $C =$ El estudiante contesta correctamente la pregunta. Así, se pregunta $P(S \mid C)$. Por el teorema de Bayes, $P(S \mid C) = P(C \mid S)P(S)/(P(C \mid S)P(S) + P(C \mid S^c)P(S^c))$
$= (1)(0.8)/((1)(0.8) + (1/4)(0.2)) = 0.941\,.$

Variables aleatorias

124. *a*) Discreta, los valores pueden ser $0, 1, 2, \ldots$ si la escala de medición son años.
b) Discreta con posibles valores $0, 1, 2, \ldots$
c) Puede ser continua y considerarse teóricamente el intervalo $[0, \infty)$.
d) Discreta con posibles valores $0, 1, 2, \ldots$ si sólo se consideran unidades monetarias.

125. *a*) Continua con valores en $(0, 1)$.
b) Discreta con valores en el conjunto $\{0, 1, 2, \ldots, 8, 9\}$.
c) Continua con valores en $(0, 1)$. d) Continua con valores en $(0, 1/10)$.

126. El rango de X es el conjunto $\{-4, -2, 0, 2, 4, 6\}$.

a) $P(X = 0) = 1/6$.
b) $P(X \in \{2, 3\}) = 1/6$.
c) $P(X \geq 0) = 2/3$.
d) $P(X < 0) = 1/3$.
e) $P(X^2 = 1) = 0$.
f) $P(2X - 4 = 0) = 1/6$.
g) $P(X^2 = 4) = 1/3$.
h) $P(|X - 2| \leq 4) = 5/6$.

127. *a*) $P(X \geq 0) = 1/2$.
b) $P(X < 0) = 1/2$.
c) $P(Y > X) = 1/2$.
d) $P(X + Y \leq 1) = 3/2 - \pi/4$.
e) $P(Z < 1/2) = 1/4$.
f) $P(1/3 < Z < 1/2) = 5/36$.

Funciones de densidad y de distribución

128. *a*) $c = 1/55$, $P(X \in \{2, 3, 4\}) = 9/55$, $P(X < 3) = 3/55$.
b) $c = 1/385$, $P(X \in \{2, 3, 4\}) = 9/385$, $P(X < 3) = 3/385$.

129. $c = 1/2\pi$, $P(X \geq \pi) = (\pi - 2)/2\pi$, $P(X \in [-\pi, 2\pi]) = 1$.

130. $c = 1/2$, $P(X \in (1, \infty)) = 1/2e$.

131. a) y b) La constante c no puede ser cero. Tampoco puede ser positiva o negativa pues en ambos casos la función de probabilidad toma valores negativos, lo cual no es permitido.

132. *a*) $P(X \geq 0) = 0.8$. *b*) $P(X < 0) = 0$. *c*) $P(X^2 = 1) = 0.7$.

133. La función de probabilidad de X^2 es

x	0	1	4	9	25
$f(x)$	0.4	0.15	0.2	0.15	0.1

La de $|X|$ es

x	0	1	2	3	5
$f(x)$	0.4	0.15	0.2	0.15	0.1

Y la de $2X - 5$ es

x	-9	-7	-5	-1	1	5
$f(x)$	0.1	0.15	0.4	0.1	0.15	0.1

134. El valor de c es 0.8. La función de probabilidad de X^2 es

x	0	4
$f(x)$	0.8	0.2

135. *a*) $k = 2$.
b) $F(x) = (x + 2)/10$ para $x \in [-2, 8]$, cero antes, uno después.
c) $P(-1 \leq X \leq 3) = 2/5$, $P(X \geq 2) = 3/5$, $P(X \leq 0) = 1/5$. d) $m = 5/2$.

136. Se trata de una variable aleatoria discreta cuya función de probabilidad es $f(1) = 1/3$ y $f(2) = 2/3$. Por lo tanto $P(X = 2) = 2/3$, y $P(1 < X < 2) = 0$.
137. La función de densidad es $f(x) = 1/(2\sqrt{x})$, para $x \in (0, 1)$. $P(X = 1/2) = 0$ y $P(X > 1/2) = F(1) - F(1/2) = 1 - 1/\sqrt{2}$.
138. *a*) $\Omega = \{(1,2), (1,3), (1,4), (2,1), (2,3), (2,4), (3,1), (3,2), (3,4), (4,1), (4,2)(4,3)\}$.
b) La función de probabilidad de X es

x	3	4	5	6	7
$f(x)$	2/12	2/12	4/12	2/12	2/12

c) $F(x) = \begin{cases} 0 & \text{si } x < 3, \\ 2/12 & \text{si } 3 \le x < 4, \\ 4/12 & \text{si } 4 \le x < 5, \\ 8/12 & \text{si } 5 \le x < 6, \\ 10/12 & \text{si } 6 \le x < 7, \\ 1 & \text{si } x \ge 7. \end{cases}$
d) $P(X \ge 6) = 1/3$, $P(3 < X \le 5) = 1/2$, $P(X = 6) = 1/6$.
139. *a*) Es función de probabilidad pues es no negativa y es tal que $f(0) + f(1) + f(2) = 1$.
b) Es función de probabilidad pues es no negativa y por el teorema del binomio se cumple que $\sum_{x=0}^{4} f(x) = (3/4 + 1/4)^4 = 1$. Esta es la función de probabilidad de la distribución bin(n, p) con $n = 4$ y $p = 3/4$.
140. Ninguna de las funciones es de densidad. La primera no integra uno, y para la segunda se tiene que $f(3/2) < 0$.
141. La función de probabilidad es $f(x) = (1/2)^{x+1}$, para $x = 0, 1, 2, \ldots$ Por lo tanto $P(0 \le X < 10) = F(9) = 1 - (1/2)^{10}$.
142. $c = 3$. $F(x) = x^2/9$ para $x \in [0, 3]$, cero antes, uno después.
143. a) $c = 2$ b) $F(x) = 1 - e^{-x^2}$, para $x > 0$.

Esperanza, varianza, momentos

144. Puede considerarse la variable aleatoria constante $X = a$. Alternativamente puede tomarse a X como aquella variable aleatoria que toma los valores $a - 1$ y $a + 1$ con idéntica probabilidad $1/2$. ¿Puede usted construir una variable aleatoria continua con la propiedad indicada?
145. a) $E(X) = 7/6$. b) $E(X) = 0$.
146. a) $E(x) = 1$. b) $E(X) = 4/3$.
147. La función es no negativa. Usando fracciones parciales se comprueba la identidad

$$\frac{1}{x(x+1)} = \frac{1}{x} - \frac{1}{x+1}.$$

Entonces la suma $\sum_{x=1}^{\infty} f(x)$ es telescópica y vale uno. Por otro lado la esperanza no existe pues la suma $\sum_{x=1}^{\infty} xf(x) = \sum_{x=1}^{\infty} \frac{1}{x+1}$ es divergente.

148. Para demostrar que esta función es de densidad es necesario recordar que la derivada de la función $h(x) = \arctan x$ es $h'(x) = 1/(1+x^2)$. Por lo tanto

$$\begin{aligned}\int_{-\infty}^{\infty} \frac{1}{\pi(1+x^2)}\, dx &= \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{d}{dx}\arctan x\, dx \\ &= \frac{1}{\pi}\arctan x|_{-\infty}^{\infty} \\ &= \frac{1}{\pi}(\pi/2 - (-\pi/2)) \\ &= 1.\end{aligned}$$

La esperanza no existe por que la integral resultante no es absolutamente convergente:

$$\begin{aligned}\int_{-\infty}^{\infty} |x|\frac{1}{\pi(1+x^2)}\, dx &= \frac{2}{\pi}\int_{0}^{\infty} \frac{x}{1+x^2}\, dx \\ &\geq \frac{2}{\pi}\int_{1}^{\infty} \frac{x}{1+x^2}\, dx \\ &\geq \frac{2}{\pi}\int_{1}^{\infty} \frac{x}{x^2+x^2}\, dx \\ &= \frac{1}{\pi}\int_{1}^{\infty} \frac{1}{x}\, dx \\ &= \infty.\end{aligned}$$

149. *a*) Cierto, la constante cero cumple tal condición.
b) Falso, las siguientes dos distribuciones son distintas y ambas tienen esperanza cero.

x	-1	0	1
$f_1(x)$	1/2	0	1/2
$f_2(x)$	1/3	1/3	1/3

c) Falso, si X es una variable aleatoria con esperanza finita y positiva, entonces $-X$ tiene esperanza negativa.
d) Cierto, por ejemplo todas las constantes son variables aleatorias con varianza cero.
e) Cierto, pues la varianza es la esperanza de una variable aleatoria no negativa.
f) Falso, dos variables aleatorias constantes distintas tienen varianza cero.

150. Tanto la esperanza como la varianza son constantes. Los resultados se siguen del hecho de que la esperanza de una constante es la constante misma y la varianza de una constante es cero.

151. *a*) $E(X) = c \times P(X = c) = c \times 1 = c$.
b) $E(X^n) = c^n \times P(X = c) = c^n \times 1 = c^n$.
c) $\text{Var}(X) = (c - E(X))^2 \times P(X = c) = (c - c)^2 \times 1 = 0$.

152. *a*) $E(X) = 3$ y $\text{Var}(X) = 24/9$.
b) Esta es la distribución geo(p) con $p = 1/2$. Usaremos resultados de sumas geométricas.

$$E(X) = \sum_{x=0}^{\infty} x(1/2)^{x+1} = \sum_{x=1}^{\infty}\sum_{y=1}^{x}(1/2)^{x+1}.$$

Intercambiando el orden de las sumas se encuentra la expresión

$$\sum_{y=1}^{\infty}\sum_{x=y}^{\infty}(1/2)^{x+1} = \sum_{y=1}^{\infty}(1/2)^{y} = 1.$$

Para la varianza encontraremos primero el segundo momento, y para ello usaremos la identidad $x^2 = x(x+1) - x$:

$$E(X^2) = \sum_{x=1}^{\infty} x^2(1/2)^{x+1} = \sum_{x=0}^{\infty} x(x+1)(1/2)^{x+1} - \sum_{x=0}^{\infty} x(1/2)^{x+1}.$$

La segunda suma acaba de ser calculada y vale uno. Para la primera suma escriba $x(x+1)$ como $2\sum_{y=1}^{x} y$. Intercambiando ahora el orden de las sumas se llega a $E(X^2) = 2\sum_{y=1}^{\infty} y \sum_{x=y}^{\infty}(1/2)^{x+1} - 1 = 2\sum_{y=1}^{\infty} y(1/2)^{y} - 1 = 2^2\sum_{y=1}^{\infty} y(1/2)^{y+1} - 1 = 4 - 1 = 3$. Por lo tanto la varianza es $\text{Var}(X) = E(X^2) - E^2(X) = 3 - 1 = 2$.

153. Ambos incisos son ciertos pues tanto $E(X)$ como $\text{Var}(X)$ son constantes.

154. *a*) La función es de densidad pues es no negativa y la integral $\int_{-\infty}^{\infty} \frac{1}{2}e^{-|x|}\,dx$ puede ser escrita como $2\int_0^{\infty} \frac{1}{2}e^{-x}\,dx = \int_0^{\infty} e^{-x}\,dx = 1$.

b) La esperanza de X es $\int_{-\infty}^{\infty} x\frac{1}{2}e^{-|x|}\,dx$. Separando la parte positiva y la negativa se obtiene $\int_0^{\infty} x\frac{1}{2}e^{-x}\,dx + \int_{-\infty}^{0} x\frac{1}{2}e^{x}\,dx$. Mediante el cambio de variable $y = -x$ se comprueba que la segunda integral es el negativo de la primera, se obtiene que la esperanza es cero.

c) El segundo momento es $\int_{-\infty}^{\infty} x^2\frac{1}{2}e^{-|x|}\,dx$. Nuevamente separando la parte positiva de la negativa se obtiene $\int_0^{\infty} x^2\frac{1}{2}e^{-x}\,dx + \int_{-\infty}^{0} x^2\frac{1}{2}e^{x}\,dx$. Ahora al hacer el cambio de variable $y = -x$ en la segunda integral se obtiene que es idéntica a la primera integral. Por lo tanto se obtiene el doble de la primera, es decir, $\int_0^{\infty} x^2e^{-x}\,dx$. Usando integración por partes puede comprobarse que esta integral vale 2.

d) Por los cálculos anteriores $\text{Var}(X) = E(X^2) - E^2(X) = 2$.

e) El n-ésimo momento es $\int_{-\infty}^{\infty} x^n\frac{1}{2}e^{-|x|}\,dx$. Nuevamente separando la parte positiva de la negativa se obtiene $\int_0^{\infty} x^n\frac{1}{2}e^{-x}\,dx + \int_{-\infty}^{0} x^n\frac{1}{2}e^{x}\,dx$. Al hacer el cambio de variable $y = -x$ en la segunda integral se obtiene que es $(-1)^{n+2}\int_0^{\infty} x^n\frac{1}{2}e^{-x}\,dx$. Cuando n es impar se trata del negativo de la primera integral, de modo que la suma es cero. Cuando n es par, es idéntica a la primera integral, y entonces el resultado es $\int_0^{\infty} x^ne^{-x}\,dx$. Aplicando repetidas veces el método de integración por partes se comprueba que esta integral vale $n!$

155. *a*) Cierto pues la esperanza es lineal.

b) Falso, pues en general $\text{Var}(cX) = c^2\,\text{Var}(X)$.

c) Falso. El lado izquierdo es $\text{Var}(X)$ y el lado derecho es siempre cero.

156. La tercera igualdad no es correcta pues la varianza en general no es lineal.

157. $E(X^n) = \theta/(\theta + n)$, para $n \geq 0$.

Cuantiles

158. La correspondiente función de distribución es $F(x) = x^2$ para $0 < x < 1$. Así, $c_{0.25}^2 = 0.25$, es decir, $c_{0.25} = 0.5$. Análogamente se calculan los otros cuantiles, $c_{0.5} = 0.707$, $c_{0.75} = 0.866$, $c_1 = 1$.
159. $c_{0.25} = -1$, $c_{0.5} = 0$, $c_{0.75} = 1$, $c_1 = 2$.
160. $c_{0.75} = 3$.
161. Si $P(X = 1) = 0.8$ y $P(X = 2) = 0.2$, entonces $c_{0.25} = c_{0.75} = 1$.

Distribución uniforme discreta

162. *a)* $E(X) = \sum_{i=1}^{n} i\,\frac{1}{n} = \frac{1}{n}\sum_{i=1}^{n} i = \frac{1}{n}\,\frac{n(n+1)}{2} = (n+1)/2.$

b) $E(X^2) = \sum_{i=1}^{n} i^2\,\frac{1}{n} = \frac{1}{n}\sum_{i=1}^{n} i^2 = \frac{1}{n}\,\frac{n(n+1)(2n+1)}{6} = (n+1)(2n+1)/6.$

c) $\mathrm{Var}(X) = E(X^2) - E^2(X) = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = (n^2-1)/12.$

163. 45/100.
164. *a)* $P(X = Y) = \sum_{x=1}^{100} P(X = x)P(Y = x) = 100(1/100)^2 = 1/100\,.$

b) $P(X \le Y) = \sum_{y=1}^{100}\sum_{x=1}^{y}(1/100)^2 = (1/100)^2\sum_{y=1}^{100} y$
$= (1/100)^2 100(101)/2 = 0.505.$

c) $P(X + Y \le 50) = \sum_{y=1}^{49}\sum_{x=1}^{50-y}(1/100)^2$
$= (1/100)^2\sum_{y=1}^{49}(50-y) = (1/100)^2(49\times 50 - 49\times 50/2)$
$= 0.1225\,.$

d) $P(X - Y \ge 10) = P(Y \le X - 10) = \sum_{x=11}^{100}\sum_{y=1}^{x-10}(1/100)^2$
$= (1/100)^2\sum_{x=11}^{100}(x-10) = (1/100)^2(100(101)/2 - 10(11)/2$
$-10(90)) = 4095/(100)^2 = 0.4095\,.$

Distribución Bernoulli

165. *a)* Es muy sencillo verificar que $f(x) \ge 0$ y $f(0) + f(1) = 1$.

b) $E(X) = 0\,(1-p) + 1\,p = p.$

c) $E(X^n) = 0^n\,(1-p) + 1^n\,p = p.$

d) $\mathrm{Var}(X) = E(X^2) - E^2(X) = p - p^2 = p(1-p).$

166. *a)* $P(X_1 = 0\,|\,X_1 + X_2 = 1) = P(X_1 = 0)P(X_2 = 1)/P(X_1 + X_2 = 1) =$
$P(X_1 = 0)\;P(X_2 = 1)/(P(X_1 = 0)P(X_2 = 1) + P(X_1 = 1)P(X_2 = 0)) =$
$1/2.$

b) $P(X_1 = 1\,|\,X_1 + X_2 = 1) = 1/2.$

c) $P(X_1 + X_2 = 2\,|\,X_1 = 1) = P(X_2 = 1)P(X_1 = 1)/P(X_1 = 1) = p.$

167. *a)* $P(Y = -1) = 1 - p$ y $P(Y = 1) = p$.

b) $E(Y) = 2E(X) - 1 = 2p - 1$ y $\mathrm{Var}(Y) = 4\;\mathrm{Var}(X) = 4p(1-p)$.

c) $E(Y^n) = \begin{cases} 2p-1 & \text{si } n \text{ es impar,} \\ 1 & \text{si } n \text{ es par.} \end{cases}$

Distribución binomial

168. *a*) Claramente $f(x) \geq 0$. Además por el teorema del binomio $\sum_{x=0}^{n} f(x) = \sum_{x=0}^{n} \binom{n}{x} p^x (1-p)^{n-x} = (p+1-p)^n = 1$.

b) La esperanza es $E(X) = \sum_{x=0}^{n} x\binom{n}{x} p^x (1-p)^{n-x}$. La suma puede empezar desde uno. Factorizando np y escribiendo $n - x = (n-1) - (x-1)$ se encuentra la expresión

$$np \sum_{x=1}^{n} \frac{(n-1)!}{((n-1)-(x-1))!(x-1)!} p^{x-1}(1-p)^{(n-1)-(x-1)},$$

lo cual es $np \sum_{x=1}^{n} \binom{n-1}{x-1} p^{x-1}(1-p)^{(n-1)-(x-1)}$. Haciendo el cambio de variable $y = x - 1$ en la suma, ésta se convierte en la suma completa de la distribución bin$(n-1, p)$ y vale uno. El resultado es entonces np.

c) Para la varianza se usa la misma técnica. Se calcula primero el segundo momento escribiendo $x^2 = x(x-1) + x$, es decir,

$$\begin{aligned} E(X^2) &= \sum_{x=0}^{n} x^2 \binom{n}{x} p^x (1-p)^{n-x} \\ &= \sum_{x=0}^{n} x(x-1) \binom{n}{x} p^x (1-p)^{n-x} + \sum_{x=0}^{n} x \binom{n}{x} p^x (1-p)^{n-x}. \end{aligned}$$

La segunda suma acaba de ser calculada y vale np. La primera suma puede empezar desde 2, y procediendo como en el caso de la esperanza puede escribirse como

$$n(n-1)p^2 \sum_{x=2}^{n} \frac{(n-2)!}{((n-2)-(x-2))!(x-2)!} p^{x-2}(1-p)^{(n-2)-(x-2)},$$

lo cual es $n(n-1)p^2 \sum_{x=2}^{n} \binom{n-2}{x-2} p^{x-2}(1-p)^{(n-2)-(x-2)}$. Haciendo el cambio de variable $y = x - 2$ en la suma, ésta se convierte en la suma completa de la distribución bin$(n-2, p)$ y vale uno. El segundo momento es entonces $n(n-1)p^2 + np$. Por lo tanto $\text{Var}(X) = E(X^2) - E^2(X) = n(n-1)p^2 + np - n^2p^2 = np(1-p)$.

169. *a*) $n = 8$ y $p = 1/2$.

b) $n = u^2/(u-v)$ y $p = 1 - v/u$, suponiendo $0 \leq v < u$ tal que la expresión para n es un entero.

170. $P(Y = x) = P(n - X = x) = P(X = n - x)$. Al substituir la función de probabilidad de X resulta que Y tiene distribución binomial con el mismo número de ensayos n, pero la probabilidad de éxito es ahora $1 - p$. Si X cuenta el número de éxitos en n ensayos Bernoulli, entonces Y cuenta el número de fracasos.

171. Desarrolle el lado derecho.

172. $\binom{6}{3}(1/2)^6$.

173. $\binom{2n}{n}(1/2)^{2n}$.

174. Recuerde que $\text{Var}(X) = npq$ y $E(X) = np$. Entonces efectivamente $npq \leq np$ pues $q \leq 1$.

175. Los tres eventos tienen la misma probabilidad de ocurrir, cada uno de ellos tiene probabilidad $1/2^6$.

176. a) 0.07463
    b) 0.5987

Distribución geométrica

177. *a*) $f(x) \geq 0$. Además $\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} p(1-p)^x = p(1/p) = 1$.
    b) Usaremos resultados de sumas geométricas. $E(X) = \sum_{x=0}^{\infty} xp(1-p)^x = \sum_{x=1}^{\infty}\sum_{y=1}^{x} p(1-p)^x$. Intercambiando el orden de las sumas se encuentra la expresión $\sum_{y=1}^{\infty}\sum_{x=y}^{\infty} p(1-p)^x = \sum_{y=1}^{\infty}(1-p)^y = (1-p)/p$.
    c) Para la varianza encontraremos primero el segundo momento, y para ello usaremos la identidad $x^2 = x(x+1)-x$. $E(X^2) = \sum_{x=1}^{\infty} x^2 p(1-p)^x = \sum_{x=0}^{\infty} x(x+1)p(1-p)^x - \sum_{x=0}^{\infty} xp(1-p)^x$. La segunda suma acaba de ser calculada y vale (1-p)/p. Para la primera suma escriba $x(x+1)$ como $2\sum_{y=1}^{x} y$. Intercambiando ahora el orden de las sumas se llega a $E(X^2) = 2\sum_{y=1}^{\infty} y \sum_{x=y}^{\infty} p(1-p)^x - (1-p)/p = 2\sum_{y=1}^{\infty} y(1-p)^y - (1-p)/p = \frac{2}{p}\sum_{y=1}^{\infty} yp(1-p)^y - (1-p)/p = 2(1-p)/p^2 - (1-p)/p$. Por lo tanto la varianza es $\text{Var}(X) = E(X^2) - E^2(X) = 2(1-p)/p^2 - (1-p)/p - (1-p)^2/p^2 = (1-p)/p^2$.

178. $P(X \geq a+b \mid X \geq a) = P(X \geq a+b)/P(X \geq a) = p\sum_{x=a+b}^{\infty}(1-p)^x/(p\sum_{x=a}^{\infty}(1-p)^x) = (1-p)^{a+b}/(1-p)^a = (1-p)^b = P(X \geq b)$.

179. $P(X \geq x) = \sum_{k=x}^{\infty} p(1-p)^k = p\sum_{k=x}^{\infty}(1-p)^k = p\frac{(1-p)^x}{1-(1-p)} = (1-p)^x$.

180. 8/3=2.66 extracciones.

181. Nos interesa el número de ensayos necesarios para obtener el primer éxito y ello corresponde a la variable $1+X$ en donde X es el número de fracasos antes del primer éxito, es decir, X tiene distribución geo(p) de acuerdo a como fue definida en este texto.
    a) En este caso $p = 1/6$ y por lo tanto $E(1+X) = 1/p = 6$.
    b) En este caso $p = 1/21$ pues hay 21 resultados distintos bajo las condiciones mencionadas. Por lo tanto, $E(1+X) = 1/p = 21$.

Distribución Poisson

182. *a*) La función $f(x)$ es no negativa y es tal que $\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1$.
    b) Para la esperanza tenemos que $E(X) = \sum_{x=0}^{\infty} x\, e^{-\lambda}\frac{\lambda^x}{x!} = \sum_{x=1}^{\infty} e^{-\lambda}\frac{\lambda^x}{(x-1)!} = \lambda\sum_{x=1}^{\infty} e^{-\lambda}\frac{\lambda^{x-1}}{(x-1)!} = \lambda$.
    c) Usando la igualdad $x^2 = x(x-1)+x$, puede encontrarse que el segundo momento es $E(X^2) = \lambda^2+\lambda$. Por lo tanto $\text{Var}(X) = E(X^2) - E^2(X) = \lambda^2+\lambda-\lambda^2 = \lambda$.

183. Desarrolle el lado derecho.

184. Sume las series $e^{\lambda} = 1+\lambda+\lambda^2/2!+\lambda^3/3!+\cdots$, y $e^{-\lambda} = 1-\lambda+\lambda^2/2!-\lambda^3/3!+\cdots$

185. a) 0.323 b) 0.593 c) Una o dos computadoras descompuestas tienen probabilidad máxima 0.2706.

186. *a)* $P(X=0) = e^{-2}\,2^0/0! = 0.135$.
b) $P(X \geq 4) = 1 - P(X \leq 3) = 1 - P(X=0) - P(X=1) - P(X=2) - P(X=3) = 1 - e^{-2}(2^0/0! + 2^1/1! + 2^2/2! + 2^3/3!) = 0.1428$.

187. Sea X el número de pólizas que presentan una reclamación. Entonces X tiene distribución bin$(550, 0.02)$. La probabilidad buscada es

$$P(X \leq 10) = \sum_{x=0}^{10} \binom{550}{x} (0.02)^x (0.98)^{550-x} = 0.458\,.$$

Sea $\lambda = np = (550)(0.02) = 11$. La aproximación Poisson es

$$P(X \leq 10) = \sum_{x=0}^{10} e^{-11}\,\frac{(11)^x}{x!} = 0.459\,.$$

Distribución binomial negativa

188. *a)* Este ejercicio no es fácil. Una forma de resolverlo es a través de la fórmula $\binom{-n}{k} = (-1)^k \binom{n+k-1}{k}$, y la expansión $(1+t)^a = \sum_{x=0}^{\infty} \binom{a}{x} t^x$ para a real y $|t| < 1$. Entonces
$\sum_{x=0}^{\infty} \binom{r+x-1}{x} p^r (1-p)^x = p^r \sum_{x=0}^{\infty} (-1)^x \binom{-r}{x} (1-p)^x$
$= p^r \sum_{x=0}^{\infty} \binom{-r}{x} (p-1)^x = p^r (1+p-1)^{-r} = 1$.
b) Para la esperanza, factorice el término $r(1-p)/p$ en la expresión $E(X) = \sum_{x=1}^{\infty} x \binom{r+x-1}{x} p^r (1-p)^x$. Observe que la suma puede empezar desde 1. La suma resultante es uno pues se trata de las probabilidades de la distribución binneg(r, p).
c) Para la varianza use la fórmula $x^2 = x(x-1) + x$ para calcular primero $E(X^2)$, separando las sumas. El procedimiento es análogo al caso de la esperanza, el resultado es $E(X^2) = r(r+1)(1-p)^2/p^2 + r(1-p)/p$. Al substituir estos resultados en la fórmula $\text{Var}(X) = E(X^2) - E^2(X)$ se obtiene $\text{Var}(X) = r(1-p)/p^2$.

189. Factorice el término $(1-p)(r+x-1)/x$ en la expresión de $P(X=x)$.

190. Sea X el número de caras hasta obtener cinco cruces. Entonces X tiene distribución binneg(r,p) con $r = 5$ y $p = 1/2$. Se usa la fórmula $E(X) = r(1-p)/p$.
a) $E(X) = 5$.
b) $E(X+5) = 10$.

Distribución hipergeométrica

191. El resultado se sigue de la fórmula $\sum_{k=0}^{r} \binom{n}{k}\binom{m}{r-k} = \binom{n+m}{r}$.

192. *a)* $P(X=0) = \dfrac{\binom{2}{0}\binom{18}{2}}{\binom{20}{2}} = 0.805$.

b) $P(X=2) = \dfrac{\binom{2}{2}\binom{18}{0}}{\binom{20}{2}} = 0.005$.

193. $P(X=1) = \dfrac{\binom{1}{1}\binom{4}{1}}{\binom{5}{2}} = 2/5\,.$

194. Sea X el número de artículos defectuosos en la muestra. Entonces X tiene distribución hipergeo(N, K, n) con $N = 1000$, $K = 50$ y $n = 20$. Se efectúa la compra cuando

$$P(X \leq 1) = P(X=0) + P(X=1) = \frac{\binom{50}{0}\binom{950}{20}}{\binom{1000}{20}} + \frac{\binom{50}{1}\binom{950}{19}}{\binom{1000}{20}} = 0.7360\,.$$

Distribución uniforme continua

195. *a*) Claramente $f(x) \geq 0$, y la integral de $f(x)$ es el área del rectángulo de base $b-a$ y altura $1/(b-a)$ que vale uno.
b) $E(X) = \int_a^b x\, 1/(b-a)\, dx = (b^2-a^2)/(2(b-a)) = (a+b)/2$.
c) $E(X^2) = \int_a^b x^2\, 1/(b-a)\, dx = (b^3-a^3)/(3(b-a)) = (a^2+ab+b^2)/3$. Por lo tanto $\text{Var}(x) = E(X^2) - E^2(X) = (a^2+ab+b^2)/3 - (a+b)^2/4 = (b-a)^2/12$.

196. $a = 1$ y $b = 7$.

197. $E(X^n) = \int_0^1 x^n\, 1\, dx = 1/(n+1)$.

198. $E(X^n) = \int_a^b x^n\, 1/(b-a)\, dx = \frac{1}{(n+1)(b-a)} x^{n+1}\big|_a^b = (b^{n+1} - a^{n+1})/((n+1)(b-a))$.

199. $E(X^n) = \int_{-1}^1 x^n\, 1/2\, dx = \frac{1}{2(n+1)}\, x^{n+1}\big|_{-1}^1$. Si n es impar, entonces $n+1$ es par y la integral se anula. Si n es par, entonces $n+1$ es impar y la integral es $1/(n+1)$.

200. X toma valores en $(0,1)$ y Y toma valores en $(-5,5)$. Para cualquier $y \in (-5,5)$, $F_Y(y) = P(Y \leq y) = P(10X - 5 \leq y) = P(X \leq (y+5)/10) = F_X((y+5)/10)$. Derivando $f_Y(y) = f_X((y+5)/10)\, 1/10$. Pero $f_X((y+5)/10)$ vale uno sólo cuando $0 < (y+5)/10 < 1$, es decir, si $-5 < y < 5$. Por lo tanto, Y tiene distribución unif$(-5,5)$.

201. La mediana en este caso es única y vale $m = (a+b)/2$ pues cumple la condición $P(X \leq m) = P(X \geq m) = 1/2$. Vea un ejercicio anterior para la demostración de que la media es también $(a+b)/2$.

Distribución exponencial

202. *a*) $f(x) \geq 0$ y $\int_0^\infty \lambda\, e^{-\lambda x}\, dx = -e^{-\lambda x}|_0^\infty = 1$.
b) $E(X) = \int_0^\infty x\, \lambda\, e^{-\lambda x}\, dx$. Usando integración por partes con $u = x$ y $dv = \lambda\, e^{-\lambda x}\, dx$, se obtiene
$E(X) = \int_0^\infty e^{-\lambda x}\, dx = (1/\lambda) \int_0^\infty \lambda\, e^{-\lambda x}\, dx = 1/\lambda$.
c) Para la varianza calcule primero el segundo momento, usando nuevamente integración por partes,

$$\begin{aligned} E(X^2) &= \int_0^\infty x^2\, \lambda\, e^{-\lambda x}\, dx = 2\int_0^\infty x\, e^{-\lambda x}\, dx \\ &= (2/\lambda) \int_0^\infty x\, \lambda\, e^{-\lambda x}\, dx \\ &= 2/\lambda^2. \text{ Por lo tanto, } \text{Var}(X) = E(X^2) - E^2(X) \\ &= 2/\lambda^2 - 1/\lambda^2 = 1/\lambda^2\,. \end{aligned}$$

203. $P(X \geq x+y \mid X \geq x) = P(X \geq x+y)/P(X \geq x) = e^{-\lambda(x+y)}/e^{-\lambda x} = e^{-\lambda y} = P(X \geq y)$.

204. $F(x)\,F(y) = (1-e^{-\lambda x})(1-e^{-\lambda y}) = 1 - e^{-\lambda x} - e^{-\lambda y} + e^{-\lambda(x+y)} = (1-e^{-\lambda x}) + (1-e^{-\lambda y}) - (1-e^{-\lambda(x+y)}) = F(x) + F(y) - F(x+y)$.

205. Suponga que el tiempo se mide en minutos. $E(X) = 10 = 1/\lambda$. Por lo tanto $\lambda = 1/10$. Entonces $P(X > 60) = 1 - P(X \leq 60) = 1 - (1 - e^{-60/10}) = e^{-6} = 0.002478$. Por lo tanto de 1000 usuarios, aproximadamente 2.47 tienen conexión mayor a una hora. a) $P(X < 10) = F(10) = 1 - e^{-10/10} = 1 - e^{-1} = 0.6321$. b) $P(10 < X < 60) = F(60) - F(10) = (1-e^{-6}) - (1-e^{-1}) = 0.3654$.

206. $P(N = n) = \int_n^{n+1} \lambda\, e^{-\lambda x}\, dx = -e^{-\lambda x}\Big|_n^{n+1} = (1-e^{-\lambda})(e^{-\lambda})^n$. Esta es la distribución geo(p) con $p = 1 - e^{-\lambda}$.

207. Como $F(x) = 1 - e^{-\lambda x}$, tenemos la ecuación $0.75 = 1 - e^{-2\lambda}$, de donde se obtiene $\lambda = \ln 2$.

Distribución gama

208. *a*) La función es no negativa, además haciendo el cambio de variable $t = \lambda x$ se obtiene
$\int_0^\infty \frac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda\, e^{-\lambda x}\, dx = \frac{1}{\Gamma(n)} \int_0^\infty t^{n-1}\, e^{-t}\, dt = \frac{\Gamma(n)}{\Gamma(n)} = 1$.

b) $E(X) = \int_0^\infty x \frac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda\, e^{-\lambda x}\, dx = \frac{\Gamma(n+1)}{\lambda\,\Gamma(n)} \int_0^\infty \frac{(\lambda x)^n}{\Gamma(n+1)} \lambda\, e^{-\lambda x}\, dx$. La integral vale uno, y usando la propiedad $\Gamma(n+1) = n\,\Gamma(n)$ se obtiene $E(X) = n/\lambda$.

c) $E(X^2) = \int_0^\infty x^2 \frac{(\lambda x)^{n-1}}{\Gamma(n)} \lambda\, e^{-\lambda x}\, dx =$
$\frac{\Gamma(n+2)}{\lambda^2\,\Gamma(n)} \int_0^\infty \frac{(\lambda x)^{n+1}}{\Gamma(n+2)} \lambda\, e^{-\lambda x}\, dx = (n+1)n/\lambda^2$. Por lo tanto,
$\text{Var}(X) = E(X^2) - E^2(X) = (n+1)n/\lambda^2 - n^2/\lambda^2 = n/\lambda^2$.

d) Análogo al inciso anterior.

209. *a*) $\Gamma(n+1) = \int_0^\infty t^n\, e^{-t}\, dt$. Use integración por partes con $u = t^n$ y $dv = e^{-t}\, dt$.

b) Por el inciso anterior, $\Gamma(2) = 1\,\Gamma(1)$, por lo tanto es suficiente demostrar que $\Gamma(1) = 1$, pero $\Gamma(1) = \int_0^\infty e^{-t}\, dt = 1$.

c) Esto es consecuencia de los dos incisos anteriores.

d) Haciendo el cambio de variable $x^2/2 = t$,

$$\Gamma(1/2) = \int_0^\infty t^{-1/2}\, e^{-t}\, dt = \sqrt{2} \int_0^\infty e^{-x^2/2} dx = \sqrt{2}\sqrt{2\pi}\,\frac{1}{2} \int_{-\infty}^\infty \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}\, dx = \sqrt{\pi}\,.$$

El último integrando es la densidad normal estándar.

Distribución beta

210. *a*) Claramente $f(x) \geq 0$, y $\int_0^1 \frac{1}{B(a,b)}\, x^{a-1}\,(1-x)^{b-1}\, dx = \frac{1}{B(a,b)} \int_0^1 x^{a-1}\,(1-x)^{b-1}\, dx = \frac{B(a,b)}{B(a,b)} = 1$.

b) $E(X) = \int_0^1 \frac{x}{B(a,b)}\, x^{a-1}\,(1-x)^{b-1}\,dx = \frac{B(a+1,b)}{B(a,b)} \int_0^1 \frac{1}{B(a+1,b)}\, x^{a}\,(1-x)^{b-1}\,dx = \frac{B(a+1,b)}{B(a,b)}$. Ahora se usa la identidad $B(a+1,b) = \frac{a}{a+b} B(a,b)$ para obtener $E(X) = a/(a+b)$.

c) $E(X^2) = \int_0^1 \frac{x^2}{B(a,b)}\, x^{a-1}\,(1-x)^{b-1}\,dx = \frac{B(a+2,b)}{B(a,b)} \int_0^1 \frac{1}{B(a+2,b)}\, x^{a+1}\,(1-x)^{b-1}\,dx = \frac{B(a+2,b)}{B(a,b)}$. En donde $B(a+2,b) = \frac{a+1}{a+b+1} B(a+1,b) = \frac{a+1}{a+b+1}\frac{a}{a+b} B(a,b)$. Por lo tanto $E(X^2) = \frac{a+1}{a+b+1}\frac{a}{a+b}$, y entonces $\text{Var}(X) = \frac{a+1}{a+b+1}\frac{a}{a+b} - \frac{a^2}{(a+b)^2} = \frac{ab}{(a+b+1)(a+b)^2}$.

211. *a)* Efectúe el cambio de variable $y = 1 - x$.

b) $B(a,1) = \int_0^1 a x^{a-1}\,dx = 1/a$.

c) $B(1,b) = \int_0^1 a(1-x)^{b-1}\,dx = 1/b$.

d) $B(a+1,b) = \int_0^1 x^a (1-x)^{b-1}\,dx$. Use integración por partes con $u = x^a$ y $dv = (1-x)^{b-1}\,dx$ para llegar a $B(a+1,b) = \frac{a}{b}\int_0^1 x^{a-1}(1-x)^b\,dx = \frac{a}{b} B(a,b+1)$.

e) Puede usarse la identidad $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(a+b)$. $B(a+1,b) = \Gamma(a+1)\Gamma(b)/\Gamma(a+b+1) = \frac{a}{a+b}\Gamma(a)\Gamma(b)/\Gamma(a+b) = \frac{a}{a+b} B(a,b)$.

f) Por el inciso anterior, $B(a,b+1) = B(b+1,a) = \frac{b}{a+b} B(b,a) = \frac{b}{a+b} B(a,b)$.

g) Efectúe el cambio de variable $x = \cos^2\theta$.

212. Simplemente substituya esos valores para los parámetros a y b en la distribución beta(a,b). Recuerde que $B(1,1) = 1$.

213. *a)* $f(x) = 1/(\pi\sqrt{x(1-x)}$, para $0 < x < 1$.

b) Efectúe el cambio de variable $x = \cos^2\theta$.

c) $E(X) = B(1+1/2,1/2)/B(1/2,1/2) = (1/2)/(1/2+1/2) = 1/2$. Para la varianza use $E(X^2) = B(2+1/2,1/2)/B(1/2,1/2)$.

214. *a)* Compruebe que $B(a,1) = 1/a$. Por lo tanto, $F(x) = (\int_0^x u^{a-1}du)/B(a,1) = x^a$, para $x \in (0,1)$.

b) Compruebe que $B(1,b) = 1/b$. Por lo tanto, $F(x) = (\int_0^x (1-u)^{b-1}du)/B(1,b) = 1-(1-x)^b$, para $x \in (0,1)$.

215. $E(X^n) = (\int_0^1 x^n\, x^{a-1}(1-x)^{b-1}dx)/B(a,b) = B(a+n,b)/B(a,b)$.

Distribución Weibull

216. La función es claramente no negativa, por lo que únicamente hace falta verificar que integra uno sobre el eje real. Haciendo el cambio de variable indicado tenemos que

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^{\infty} e^{-u(x)}\,u'(x)\,dx = -e^{-u(x)}\Big|_0^{\infty} = 1.$$

217. Haciendo el cambio de variable indicado, para valores positivos de x,

$$F(x) = \int_0^x e^{-u(t)}\,u'(t)\,dt = -e^{-u(t)}\Big|_0^x = 1 - e^{-u(x)} = 1 - e^{-(\lambda x)^\alpha}.$$

218. $P(X > 5) = 1 - F(5) = e^{-\sqrt{5}} = 0.10697$. Se espera que el componente funcione bien en promedio $E(X) = 2$ años.

Distribución normal

219. *a*) Este ejercicio no es fácil pero puede resolverse por varios métodos. El siguiente método hace uso de coordenadas polares. Sea $I = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}\, dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$. Entonces $I^2 = (\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx)\,(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy)$ $= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{1}{2\pi} e^{-(x^2+y^2)/2}\, dx\, dy = \int_0^{2\pi}\int_0^{\infty} \frac{1}{2\pi} e^{-r^2/2} r\, dr\, d\theta$, en donde la última integral se obtiene después del cambio de variable a coordenadas polares $(x, y) = (r\cos\theta, r\,\text{sen}\,\theta)$. Entonces $I^2 = \int_0^{\infty} e^{-r^2/2} r\, dr = 1$.

b) $E(X) = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}\, dx$. Haciendo el cambio de variable $u = (x-\mu)/\sigma$ se encuentra que $E(X) = \mu + \sigma E(Z)$, en donde $Z \sim \text{N}(0, 1)$. Por lo tanto es suficiente demostrar que $E(Z) = 0$, pero ello es fácil pues el integrando de $E(Z) = \int_{-\infty}^{\infty} u \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, du$ es una derivada excepto por el signo. Por lo tanto $E(X) = \mu$.

c) Para la varianza considere el segundo momento $E(X^2) = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/2\sigma^2}\, dx$. Efectúe nuevamente el cambio de variable $y = (x-\mu)/\sigma$ y encuentre que $E(X^2) = \mu^2 + 2\mu E(Z)$ $+\sigma^2 E(Z^2)$. Resta encontrar $E(Z^2) = \int_{-\infty}^{\infty} y^2 \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$. Esta integral puede realizarse por partes con $u = y$ y $dv = y \frac{1}{\sqrt{2\pi}} e^{-y^2/2}\, dy$. El resultado es 1. Por lo tanto $E(X^2) = \mu^2 + \sigma^2$, y entonces $\text{Var}(X) = (\mu^2 + \sigma^2) - \mu^2 = \sigma^2$.

220. Suponga $X \sim \text{N}(\mu, \sigma^2)$. Sea $Z = (X-\mu)/\sigma$. Entonces $F_Z(u) = P(Z \leq u) = P(\frac{X-\mu}{\sigma} \leq u) = P(X \leq \mu + u\sigma) = F_X(\mu + u\sigma)$. Derivando respecto a u, $f_Z(u) = f_X(\mu + u\sigma)\sigma$. Substituya la expresión para $f_X(x)$ y encuentre la densidad normal estándar. El procedimiento es reversible. Suponga $Z \sim \text{N}(0, 1)$. Sea $X = \mu + \sigma Z$. Entonces $F_X(x) = P(X \leq x) = P(\mu + \sigma Z \leq x) = P(Z \leq (x-\mu)/\sigma) = F_Z((x-\mu)/\sigma)$. Derivando respecto a x, $f_X(x) = f_Z((x-\mu)/\sigma)/\sigma$. Substituya la expresión de $f_Z(x)$ y encuentre la densidad $\text{N}(\mu, \sigma^2)$.

221. *a*) $P(X \geq 10) = 1/2$.
b) $P(X < 0) = P(Z < -2) = 0.0228$.
c) $P(0 < X \leq 10) = P(-2 < Z \leq 0) = 0.4772$.
d) $P(X \geq 20) = P(Z \geq 2) = 0.0228$.
e) $P(-20 < X \leq 10) = P(-6 < Z \leq 0) = 1/2$.

222. *a*) $P(X \leq 10) = P(Z \leq 1) = 0.8413$.
b) $P(X > 0) = 1/2$.
c) $P(0 < X \leq 40) = P(0 < Z \leq 4) = 1/2$.
d) $P(X \geq 30) = P(Z \geq 3) = 0.0013$.
e) $P(-10 < X \leq 10) = P(-1 < Z \leq 1) = 0.6826$.

223. a) $x = 1.115$. b) $x = -1.375$.

224. Suponga $a > 0$. Entonces $F_Y(y) = P(Y \leq y) = P(aX + b \leq y) = P(X \leq (y-b)/a) = F_X((y-b)/a)$. Derivando, $f_Y(y) = f_X((y-b)/a)/a$. Substituya ahora en esta expresión la función de densidad $\text{N}(\mu, \sigma^2)$ y encuentre la función de densidad normal ahora con media $a\mu + b$ y varianza $a^2\sigma^2$. Haga un análisis semejante

para el caso $a < 0$, tenga cuidado al dividir por tal cantidad. Los nuevos parámetros son idénticos al caso anterior.

225. Sea $Y = -X$. Entonces $F_Y(y) = P(Y \leq y) = P(-X \leq y) = P(X \geq -y) = 1 - P(X < -y) = 1 - F_X(-y)$. Derivando, $f_Y(y) = f_X(-y)$. Substituya ahora la función de densidad de X evaluada en $-y$ y compruebe que se trata de la densidad normal con media $-\mu$ y varianza σ^2.

226. Para $x > 0$, $F_{X^2}(x) = P(X^2 \leq x) = P(-\sqrt{x} \leq X \leq \sqrt{x}) = F_X(\sqrt{x}) - F_X(-\sqrt{x})$. Derivando, $f_{X_2}(x) = f_X(\sqrt{x})\frac{1}{2\sqrt{x}} + f_X(-\sqrt{x})\frac{1}{2\sqrt{x}}$
$= f_X(\sqrt{x})\frac{1}{\sqrt{x}}$. Ahora substituya la expresión para $f_X(x)$ y encuentre la función de densidad de la distribución $\chi^2(1)$.

227. Para $y > 0$, $F_Y(y) = P(Y \leq y) = P(|X| \leq y) = P(-y \leq X \leq y) = F_X(y) - F_X(-y)$. Derivando, $f_Y(y) = f_X(y) + f_X(-y) = 2f_X(y)$. Ahora substituya la función de densidad de X.

228. Usando la tabla de la distribución normal estándar se encuentra que $c_{0.25} = 0.501$, $c_{0.50} = 2$, y $c_{0.75} = 3.498$.

229. Sea X el llenado de un vaso cualquiera. Suponga $X \sim \text{N}(320, 100)$. Entonces
   a) $P(X > 310) = P(Z > 1) = 0.1587$.
   b) $P(290 < X < 305) = P(-1 < Z < 1/2) = 0.5328$.
   c) $P(X > 320) = P(Z > 2) = 0.0228$.
   d) $P(X < 270) = P(Z < -3) = 0.0013$. De mil clientes, 1.3 reclamarán por vasos servidos con 270 ml. o menos.

Distribución ji cuadrada

230. *a*) La función es no negativa. Efectúe el cambio de variable $t = x/2$ dentro de la integral. La integral resultante es la función gama evaluada en $n/2$.
   b) En estos cálculos es conveniente hacer el cambio de variable $t = x/2$ en cada una de las integrales. Para la esperanza, el exponente de t es $n/2$, que puede ser escrito como $(n/2) + 1 - 1$. Reconstruya la integral como la función gama evaluada en $(n/2) + 1$. Después simplifique usando la propiedad $\Gamma(n + 1) = n\,\Gamma(n)$.
   c) Calcule primero el segundo momento, el exponente de t es $(n/2) + 1$, que puede ser escrito como $(n/2) + 2 - 1$. Ahora reconstruya la integral resultante como la función gama evaluada en $(n/2) + 2$. Después simplifique.
   d) El cálculo del m-ésimo momento sigue las mismas líneas del inciso anterior.

231. Simplemente substituya el valor $n = 2$ en la densidad $\chi^2(n)$. Recuerde que $\Gamma(1) = 1$.

Distribución t

232. *a*) La función es no negativa. Para comprobar que esta función integra uno efectúe el cambio de variable $1 - u = (1 + x^2/n)^{-1}$ y reconstruya la integral de $B(1/2, n/2)$.
   b) La integral $E(X)$ es absolutamente convergente para $n > 1$ y el integrando es una función impar, por lo tanto la integral es cero.
   c) Dado el resultado del inciso anterior, la varianza es el segundo momento. Efectúe el cambio de variable $1 - u = (1 + x^2/n)^{-1}$ en la integral $E(X^2)$, y reconstruya la integral de $B(3/2, (n-2)/2)$, después simplifique.

233. *a*) $\int_{-\infty}^{1} \frac{dx}{1+x^2} = 3\pi/4$.
b) $\int_{-\infty}^{9} \frac{dx}{(12+x^2)^2} = \frac{\sqrt{3}}{9 \cdot 2} \frac{99}{100}\pi = 0.299$.

Simulación de v.a.s discretas

234. Use el procedimiento explicado en la página 79.
235. Use el procedimiento explicado en la página 79 y la ley de los grandes números.
236. Puede usarse el método general de la página 79, sin embargo dado que X es el número de fracasos antes de obtener el primer éxito en una sucesión de ensayos independientes Bernoulli, pueden generarse valores $y_1, y_2, \ldots$ de la distribución Ber(p) y definir un valor observado de X como el número de ceros obtenidos hasta que aparece por primera vez un uno. Así, cada sucesión generada $y_1, y_2, \ldots$ determina una observación para X.
237. Puesto que X es el número de fracasos antes de obtener el r-ésimo éxito en una sucesión de ensayos independientes Bernoulli, pueden entonces generarse valores y_1, $y_2, \ldots$ de la distribución Ber(p) y definir un valor observado de X como el número de ceros obtenidos hasta que aparece el r-ésimo uno. Así, cada sucesión generada y_1, $y_2, \ldots$ determina una observación para X.

Simulación de v.a.s continuas

238. Genere valores de la variable X, eleve al cuadrado cada uno de estos valores y efectúe el promedio aritmético de estas cantidades. El resultado es una estimación de $E(X^2)$ y debe ser cercano a
$$E(X^2) = \int_0^1 x^2\, dx = \frac{1}{3}.$$
239. Para cada pareja de valores x, y que genere, suponga que se observa un éxito si $y > x$ en caso contrario es un fracaso. Por la ley de los grandes números, el cociente entre el número de éxitos entre el número de ensayos debe ser cercano a $1/3$ cuando el número de ensayos crece a infinito.
240. Si x es un valor de X, entonces $y = \frac{1}{\sqrt{2\pi}}\, e^{-x^2/2}$ es un valor de la variable Y.

Variables y tipos de datos

241. *a*) Cualitativa, escala ordinal.
b) Cuantitativa, discreta, escala de razón.
c) Cualitativa, escala ordinal.
d) Cuantitativa, discreta, escala de razón.
e) Cualitativa, escala ordinal.
f) Cualitativa, escala nominal.
g) Cualitativa, escala nominal.
h) Cualitativa, escala nominal.
i) Cuantitativa, discreta, escala de razón.

Estadística descriptiva

242. Aplique las fórmulas respectivas.
243. a) $\bar{u} = \frac{1}{n}\sum_{i=1}^{n} u_i = \frac{1}{n}\sum_{i=1}^{n}(ax_i + b) = \frac{1}{n}(a\sum_{i=1}^{n} x_i + nb) = a\bar{x} + b$.
b) $s^2(u_1, \ldots, u_n) = \frac{1}{n-1}\sum_{i=1}^{n}(u_i - \bar{u})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(ax_i + b - (a\bar{x} + b))^2$
$= \frac{1}{n-1}\sum_{i=1}^{n}(a(x_i - \bar{x}))^2 = a^2\, s^2(x_1, \ldots, x_n)$.
244. Aplique las fórmulas respectivas.
245. Si los datos se generan de acuerdo a la distribución uniforme sobre el intervalo $(0, 1)$, entonces la media y mediana convergen a $1/2$. La varianza debe converger a $1/12 = 0.8\overline{3}$, la desviación estándar a 0.2886751 y el rango debe converger a 1. El conjunto de datos generado será multimodal y se acercará al conjunto de todos los posibles valores del intervalo unitario susceptibles de ser generados cuando se considere que un valor pertenece al conjunto modal si su frecuencia dista de la frecuencia modal observada real en menos de algún $\epsilon > 0$.
246. *a*) $\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - n\bar{x} = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0$.
b) $\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(x_i^2 - 2x_i\bar{x} + \bar{x}^2) = \sum_{i=1}^{n} x_i^2 - 2\bar{x}\sum_{i=1}^{n} x_i + n\bar{x}^2 = \sum_{i=1}^{n} x_i^2 - 2n\bar{x}^2 + n\bar{x}^2 = \sum_{i=1}^{n} x_i^2 - n\bar{x}^2$.
c) Idéntico al inciso anterior. Empiece desarrollando s^2.
247. Derive la función $\bar{x}(c)$ respecto de la variable c, iguale a cero y encuentre que $c = \bar{x}$. Compruebe que la segunda derivada es negativa.
248. Calcule la mitad de cada intervalo como valor representativo. Suponga que ese valor medio se observa tantas veces como indica la frecuencia en la tabla. Calcule ahora las cantidades solicitadas a partir de este conjunto aproximado de observaciones. Para elaborar el histograma tome como base cada intervalo y como altura la frecuencia.
249. Siga el procedimiento explicado en la solución anterior.

Vectores aleatorios

250. (X, Y) tiene distribución uniforme en el conjunto que consta de los doce elementos: $(1, 2)$,$(1, 3)$, $(1, 4)$,$(2, 1)$,$(2, 3)$,$(2, 4)$,$(3, 1)$,$(3, 2)$,$(3, 4)$, $(4, 1)$,$(4, 2)$,$(4, 3)$. Tanto X como Y tienen distribución uniforme en el conjunto $\{1, 2, 3, 4\}$.
251. *a*) Se trata de una función no negativa, cuyos valores suman uno.
b) La función de densidad marginal de X es $f_1(1) = 0.8$ y $f_1(2) = 0.2$. La de la variable Y es $f_2(-1) = 0.35$ y $f_2(0) = 0.65$.
c) La función de distribución marginal de X es $F_1(x) = 0$ para $-\infty < x < 1$, $F_1(x) = 0.8$ para $1 \leq x < 2$, $F_1(x) = 1$ para $x \geq 2$. Análogamente se construye la función de distribución de Y.
d) No lo son pues por ejemplo $P(X = 1, Y = -1) = 0.3$ que es distinto a $P(X = 1)P(Y = -1) = 0.8 \times 0.35 = 0.28$.
252. *a*) Nuevamente se trata de una función no negativa, cuyos valores suman uno.
b) La función de densidad marginal de X es $f_1(1) = 0.2$ y $f_1(2) = 0.8$. La de la variable Y es $f_2(0) = 0.9$ y $f_2(1) = 0.1$.
c) La función de distribución marginal de X es $F_1(x) = 0$ para $-\infty < x < 1$, $F_1(x) = 0.2$ para $1 \leq x < 2$, $F_1(x) = 1$ para $x \geq 2$. Análogamente se construye la función de distribución de Y.

d) No lo son pues por ejemplo $P(X = 1, Y = 1) = 0$ que es distinto a $P(X = 1)$ $P(Y = 1) = 0.2 \times 0.1 = 0.02$.

253. *a*) Tal función toma el valor $1/4$ sobre el cuadrado $(-1, 1) \times (-1, 1)$ y vale cero fuera de él. Por lo tanto se trata de una función no negativa que integra uno.
b) Las funciones de densidad marginales coinciden en la función $f_1(x) = 1/2$ para x en $(-1, 1)$, con valor nulo fuera de este intervalo.
c) Las funciones de distribución marginales son por tanto coincidentes y son $F_1(x) = 0$ para $-\infty < x < -1$, $F_1(x) = (x + 1)/2$ para $-1 \leq x < -1$, $F_1(x) = 1$ para $x \geq 1$.
d) Estas variables aleatorias son independientes pues se cumple la identidad $f(x, y) = f_1(x) f_2(y)$, para cualquier valor de x y y.

254. *a*) La función es no negativa y puede comprobarse que integra uno sobre $\mathbb{R}^2$.
b) La distribución marginal de esta variable es $\exp(\lambda)$ con $\lambda = 1$.
c) Idéntico al inciso anterior.
d) Las variables son independientes pues se cumple la identidad

$$f(x, y) = f_1(x) f_2(y)\,,$$

para cualquier valor de x y y.

Covarianza

255. $(X - E(X))(Y - E(Y)) = XY - XE(Y) - YE(X) + E(X)E(Y)$. Ahora aplique esperanza y simplifique, recuerde que $E(X)$ y $E(Y)$ son constantes.

256. *a*) $\text{Cov}(cX, Y) = E[(cX - E(cX))(Y - E(Y))] = E[c(X - E(X))(Y - E(Y))] = c\ \text{Cov}(X, Y)$.
b) $(X_1 + X_2 - E(X_1 + X_2)) = (X_1 + X_2 - E(X_1) - E(X_2)) = [X_1 - E(X_1)] + [X_2 - E(X_2)]$. Introduzca este cálculo en la definición para $\text{Cov}(X_1 + X_2, Y)$ y simplifique.

257. $E(X) = E(Y) = 0$. Compruebe además que $E(XY) = 1/2$.

258. $E(X) = E(Y) = 2/3$, $E(XY) = (2/3)^2$. O bien observe que X y Y son independientes.

Coeficiente de correlación

259. Puede comprobarse que las variables son independientes y ello es razón suficiente para que el coeficiente de correlación se anule.

260. a) $\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = E(X^3) - E(X)E(X^2)$. El primer y tercer momento se anulan, de modo que la covarianza es cero.
b) X y Y no son independientes pues, por ejemplo, $P(X = 0, Y = 0) \neq P(X = 0)P(Y = 0)$.

261. Se hace uso de la propiedad que establece que la esperanza del producto de dos variables aleatorias es el producto de las esperanzas:
a) $\text{Cov}(X_1, X_1 + \cdots + X_n) = E(X_1(X_1 + \cdots + X_n)) - E(X_1)E(X_1 + \cdots + X_n) = p + (n - 1)p - np^2 = p(1 - p)$.
b) $\rho(X_1, X_1 + \cdots + X_n) = p(1 - p)/\sqrt{np^2(1 - p)^2} = 1/\sqrt{n}$.

Muestras aleatorias y estadísticas

262. X tiene distribución bin$(3, p)$. Por lo tanto, la media es $3p$ y la varianza es $3p(1-p)$.
263. a) si. b) no. c) si. d) no. e) si. f) si.
264. La distribución es $\mathrm{N}(n\mu, n\sigma^2)$, de donde se obtienen la media y la varianza.
265. Aplicando la fórmula para el coeficiente de correlación muestral de la página 110 se obtiene $\rho(\{x\},\{y\}) = 0.8545$. Esto significa que los datos muestran que existe una buena dependencia lineal directa entre las calificaciones de una materia y la otra.

Estimación puntual

266. La esperanza es p y la varianza es $p(1-p)/n$.
267. *a)* $E(\hat{\lambda_1}) = n/\lambda$, $\mathrm{Var}(\hat{\lambda_1}) = n/\lambda^2$.
b) $E(\hat{\lambda_2}) = 1/\lambda$, $\mathrm{Var}(\hat{\lambda_2}) = 1/(2\lambda^2)$.
c) $E(\hat{\lambda_3}) = 1/\lambda$, $\mathrm{Var}(\hat{\lambda_3}) = 1/(n\lambda^2)$.

Método de momentos

268. El método de momentos en este caso establece que $\hat{a} = 2\bar{X}$. Sin embargo, dado que algunas de las observaciones pueden ser mayores a este valor, se propone como estimador $\hat{a} = \max\{2\bar{X}, X_{(n)}\}$, en donde $X_{(n)}$ es el máximo de $X_1, X_2, \ldots, X_n$.
269. $\hat{a} = \bar{X} - \sqrt{3S^2}$ y $\hat{b} = \bar{X} + \sqrt{3S^2}$. Sin embargo, si alguna observación es mayor a $\hat{b}$ se toma $\hat{b} = \max\{\bar{X} + \sqrt{3S^2}, X_{(n)}\}$, en donde $X_{(n)}$ es el máximo de $X_1, X_2, \ldots, X_n$. ¿Cómo debe tomarse $\hat{a}$ de manera general?
270. *a)* $\hat{a} = \bar{x} = 20.36$ y $\hat{b} = \sqrt{\frac{3}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2} = 1.48$.
b) $P(X > 21) = 0.283$.
271. *a)* $\hat{\lambda} = 1/\bar{x} = 1/17.83 = 0.056$.
b) $E(X) = 1/\hat{\lambda} = 17.83$.
272. $\hat{\lambda} = \bar{X}$.
273. *a)* $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 0.1325$, en donde $\bar{x} = -0.05$.
b) $P(|V| > 0.2) = P(|Z| > 0.2/0.364) = 2(1 - \Phi(0.549)) = 0.5824$.
274. La esperanza de X es cero, de modo que igualar el primer momento de la distribución con el primer momento muestral no produce una ecuación para θ. Igualando el segundo momento de la distribución, que es θ, con el segundo momento muestral lleva a la estimación $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 4/7$.
275. La esperanza de la densidad $f(x)$ es $\theta\mu_1 + (1-\theta)\mu_2$. Al igualar este valor a la media muestral $\bar{x}$ se obtiene el estimador $\hat{\theta} = (\bar{x} - \mu_2)/(\mu_1 - \mu_2)$.

Método de máxima verosimilitud

276. La función de verosimilitud está dada por $L(p) = p^{n\bar{x}}(1-p)^{n(1-\bar{x})}$. Tomando logaritmo, derivando, e igualando a cero se encuentra que $p = \bar{x}$.
277. $\hat{\lambda} = \bar{X}$.
278. $\hat{p} = n/\bar{X}$.
279. Derivando el logaritmo de la función de verosimilitud e igualando a cero se encuentra que $\hat{\theta} = -n/\sum_{i=1}^{n}\ln x_i$.

280. En este caso el máximo de la función de verosimilitud no existe. Ninguno de los valores observados $x_1, x_2, \ldots, x_n$ puede ser mayor a b, ni menor a a. Es por ello que se propone $a = \text{mín}\{x_i\}$ y $b = \text{máx}\{x_i\}$.

281. El parámetro n es un número entero positivo que no puede ser menor a ninguna de las observaciones $x_1, x_2, \ldots, x_m$. Es por ello que se toma $n = \text{máx}\{x_1, x_2, \ldots, x_m\}$. Conociendo entonces n se calcula la función de verosimilitud como función de p, se toma logaritmo, se deriva, se iguala a cero y se encuentra que $p = \bar{x}/n$ en donde $\bar{x} = (x_1 + x_2 + \cdots + x_m)/m$.

282. Derivando el logaritmo de la función de verosimilitud respecto de λ e igualando a cero se encuentra que $\hat{\lambda} = (n/\sum_{i=1}^{n} x_i^{\alpha})^{1/\alpha}$.

283. *a*) El procedimiento de derivar e igualar a cero no funciona pues la función de verosimilitud crece conforme el parámetro a crece. Las observaciones x_1, x_2, x_3, x_4 deben ser todas mayor o igual a a, así, el valor máximo que podemos tomar para a es $\hat{a} = \text{mín}\{x_1, x_2, x_3, x_4\} = 9.3$ años.
b) $E(X) = 3\hat{a}/2 = 13.95$ años.

284. Derivando el logaritmo de la función de verosimilitud respecto de λ e igualando a cero se encuentra que $\hat{\lambda} = n/\bar{x}$, en donde $\bar{x} = (x_1 + \cdots + x_m)/m$.

285. *a*) La función es no negativa y $\int_0^{\theta} 2x/\theta^2 dx = (x^2/\theta^2)\Big|_0^{\theta} = 1$.
b) La función de verosimilitud crece sin límite conforme el parámetro θ se acerca a cero. Por otro lado, todas las observaciones deben ser menores a θ, así, el valor más pequeño que podemos tomar para θ es $\hat{\theta} = \text{máx}\{1.7, 1.5, 3, 2.1, 2.5\} = 3$.
c) $E(X) = 2$.

286. *a*) $f(x) \geq 0$ y $\int_{\theta}^{\infty} e^{-(x-\theta)}\, dx = -e^{-(x-\theta)}\Big|_{\theta}^{\infty} = 1$.
b) La función de verosimilitud es $L(\theta) = e^{n(\theta - \bar{X})}$, la cual crece conforme θ crece. Así, conviene tomar el valor más grande para θ pero de tal forma que θ sea siempre menor o igual a cada una de las observaciones, es decir, $\theta \leq X_1, X_2, \ldots, X_n$. Esto lleva a tomar $\hat{\theta} = \text{mín}\{X_1, X_2, \ldots, X_n\}$.
c) $\hat{\theta} = 5$.
d) $E(X) = \hat{\theta} + 1 = 6$.

Insesgamiento

287. *a*) $E(\hat{\theta}) = E(a\hat{\theta}_1 + (1-a)\hat{\theta}_2) = aE(\hat{\theta}_1) + (1-a)E(\hat{\theta}_2) = a\theta + (1-a)\theta = \theta$.
b) $a + b = 1$.

288. a) no. b) si. c) no.

289. *a*) $E(\hat{\sigma}^2) = \dfrac{1}{n}\sum_{i=1}^{n} E(X_i - \mu)^2 = \dfrac{1}{n}\sum_{i=1}^{n} \text{Var}(X_i) = \dfrac{1}{n}\sum_{i=1}^{n} \sigma^2 = \sigma^2$.
b) Siga la sugerencia del Ejemplo 2.39 en la página 118.
c) $E(\hat{\sigma}^2) = \dfrac{1}{2(n-1)}\sum_{i=1}^{n-1} E(X_{i+1} - X_i)^2$

$= \dfrac{1}{2(n-1)}\sum_{i=1}^{n-1}(E(X_{i+1}^2) - 2E(X_{i+1})E(X_i) + E(X_i^2))$

$$= \frac{1}{2(n-1)} \sum_{i=1}^{n-1} (\sigma^2 + \mu^2 - 2\mu^2 + \sigma^2 + \mu^2) = \frac{1}{2(n-1)} \sum_{i=1}^{n-1} 2\sigma^2 = \sigma^2.$$

290. *a*) Siga la sugerencia de la página 118.
b) Sugerencia: $\frac{(n-1)}{\sigma^2} S^2 \sim \chi^2(n-1)$.

291. $E(\hat{\theta}) = E(X_1^2) = \theta$. Para el siguiente cálculo es necesario saber que el cuarto momento de una distribución $N(0, \sigma^2)$ es $3(\sigma^2)^2 = 3\sigma^4$.
$\text{Var}(\hat{\theta}) = \frac{1}{n} \text{Var}(X_1^2) = \frac{1}{n}(E(X_1^4) - E^2(X_1^2)) = \frac{1}{n}(3\theta^2 - \theta^2) = \frac{2}{n}\theta^2$.

292. $E(\bar{X}) = \theta\mu_1 + (1-\theta)\mu_2$, de donde se obtiene que $E(\hat{\theta}) = \theta$.

Estimación por intervalos

293. (1541.68, 1558.31).

294. (2223.70, 2259.39).

295. $n \geq 1075$ aproximadamente.

296. La media muestral es $\bar{x} = 111.1$. Suponiendo un modelo normal y con $z_{\alpha/2} = z_{0.025} = 1.96$, tenemos que el intervalo para la media es

$$\begin{aligned} (\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}, \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}) &= (111.1 - 1.96\frac{20}{\sqrt{10}}, 111.1 + 1.96\frac{20}{\sqrt{10}}) \\ &= (98.70, 123.49). \end{aligned}$$

Pruebas de hipótesis

297. De 38 a 62 cruces no se rechaza, en caso contrario se rechaza.

298. $\alpha = 0.1$, $\beta(\theta_1) = 0.9/\theta_1$.

Pruebas para la media de una distribución normal con varianza conocida

299. Se acepta H_0.

300. Se rechaza H_0.

301. Cuando $n \to \infty$, el cociente $\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}}$ diverge a infinito o menos infinito dependiendo del signo de la diferencia $\mu_0 - \mu_1$.

Pruebas para la media de una distribución normal con varianza desconocida

302. $\alpha = 0.05$, $t_{0.025,99} = 1.960$ y $t = (\bar{x} - \mu_0)/(s/\sqrt{n}) = 6.57$. Por lo tanto se rechaza la hipótesis H_0, es decir, la nueva máquina parece estar llenando cantidades distintas de 600ml.

303. $\bar{x} = 19.25$, $s^2 = 1.251$, $\alpha = 0.10$, $t_{0.10,9} = 1.383$ y $t = (\bar{x} - \mu_0)/(s/\sqrt{n}) = -2.11990$. Por lo tanto, $t \leq -t_{\alpha,n-1}$, y entonces se rechaza la hipótesis H_0. Los peces parecen estar disminuyendo en tamaño.

304. $\bar{x} = 65.153$, $s^2 = 347.641$, $\alpha = 0.10$, $t_{0.10,12} = 1.356$ y $t = (\bar{x} - \mu_0)/(s/\sqrt{n}) = 0.029$. Por lo tanto, no se cumple que $t \geq t_{\alpha,n-1}$ y no se rechaza la hipótesis H_0, es decir, no parece que el curso de preparación tenga algún efecto positivo en el resultado del examen.

Apéndice C

Formulario

Notación

$\mathbb{N}$ Conjunto de números naturales $1, 2, 3, \ldots$
$\mathbb{Z}$ Conjunto de números enteros $0, \pm 1, \pm 2, \pm 3, \ldots$
$\mathbb{Q}$ Conjunto de números racionales a/b en donde $a, b \in \mathbb{Z}$ con $b \neq 0$.
$\mathbb{R}$ Conjunto de números reales.

El alfabeto griego

A α	alfa	I ι	iota	P ρ, ϱ	ro
B β	beta	K κ	kapa	Σ σ, ς	sigma
Γ γ	gama	Λ λ	lambda	T τ	tau
Δ δ	delta	M μ	mu	Υ υ	upsilon
E ϵ, ε	epsilon	N ν	nu	Φ ϕ, φ	fi
Z ζ	zeta	Ξ ξ	xi	X χ	ji
H η	eta	O o	omicron	Ψ ψ	psi
Θ θ, ϑ	teta	Π π	pi	Ω ω	omega

Exponentes

a) $x^1 = x$.
b) $x^0 = 1, \quad x \neq 0$.
c) $x^{-1} = \dfrac{1}{x}, \quad x \neq 0$.
d) $x^n\, x^m = x^{n+m}$.
e) $\dfrac{x^n}{x^m} = x^{n-m}$.
f) $(x^n)^m = x^{nm}$.

g) $(xy)^n = x^n\, y^n$.
h) $\left(\dfrac{x}{y}\right)^n = \dfrac{x^n}{y^n}$.
i) $x^{-n} = \dfrac{1}{x^n}, \quad x \neq 0$.
j) $x^{m/n} = \sqrt[n]{x^m}$.

Logaritmos

a) $\log xy = \log x + \log y$.
b) $\log \frac{x}{y} = \log x - \log y$.
c) $\log x^n = n \log x$.
d) $\log \sqrt[n]{x} = \log x^{1/n} = \dfrac{1}{n} \log x$.
e) $\log 1 = 0$.
f) $\log_a a = 1, \quad a > 0$.
g) Cambio de base: $a > 0, b > 0$,

$$\log_b x = (\log_b a)\ \log_a x.$$

Sumas

a) $\displaystyle\sum_{k=m}^{n} x_k = x_m + x_{m+1} + \cdots + x_n, \qquad m \leq n.$
b) $\displaystyle\sum_{k=1}^{n} c = nc, \quad c$ constante.
c) $\displaystyle\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$.
d) $\displaystyle\sum_{k=1}^{n} k^2 = \frac{n(n+1)(2n+1)}{2}$.
e) $\displaystyle\sum_{k=1}^{n} k^3 = \left[\frac{n(n+1)(2n+1)}{2}\right]^2$.
f) $\displaystyle\sum_{k=m}^{n} a^k = \frac{a^m - a^{n+1}}{1-a}, \quad a \neq 1$.
g) $\displaystyle\sum_{k=0}^{\infty} \frac{x^k}{k!} = e^x, \quad x \in \mathbb{R}$.
h) $\displaystyle\sum_{k=0}^{n} \binom{n}{k} a^k\, b^{n-k} = (a+b)^n, \quad a, b \in \mathbb{R}, n \in \mathbb{N}$.

Derivadas

a) $\dfrac{d}{dx}\, c = 0, \quad c$ constante.

b) $\dfrac{d}{dx}\, x = 1.$

c) $\dfrac{d}{dx}\, x^n = n x^{n-1}.$

d) $\dfrac{d}{dx}\, e^x = e^x.$

e) $\dfrac{d}{dx}\, \ln x = \dfrac{1}{x}.$

f) $\dfrac{d}{dx}\, \log_a x = (\log_a e)\, \dfrac{1}{x}, \quad a > 0.$

g) $\dfrac{d}{dx}\, \operatorname{sen} x = \cos x.$

h) $\dfrac{d}{dx}\, \cos x = -\operatorname{sen} x.$

i) $\dfrac{d}{dx}\, [f(x) \pm g(x)] = f'(x) \pm g'(x).$

j) $\dfrac{d}{dx}\, [f(x)\, g(x)] = f(x)\, g'(x) + f'(x)\, g(x).$

k) $\dfrac{d}{dx}\, \dfrac{f(x)}{g(x)} = \dfrac{g(x) f'(x) - f(x) g'(x)}{g^2(x)}.$

l) $\dfrac{d}{dx}\, f(g(x)) = f'(g(x))\, g'(x).$ (Regla de la cadena)

Integrales

a) $\displaystyle\int df(x) = \int f'(x)\, dx = f(x) + c.$

b) $\displaystyle\int c\, dx = c \int dx, \quad c$ constante.

c) $\displaystyle\int x^n\, dx = \frac{x^{n+1}}{n+1} + c, \quad n \neq -1.$

d) $\displaystyle\int \frac{dx}{x} = \ln x + c.$

e) $\displaystyle\int e^{ax}\, dx = \frac{1}{a}\, e^{ax} + c.$

f) $\displaystyle\int \ln x\, dx = x \ln x - x + c.$

g) $\displaystyle\int \operatorname{sen} x\, dx = -\cos x + c.$

h) $\displaystyle\int \cos x\, dx = \operatorname{sen} x + c.$

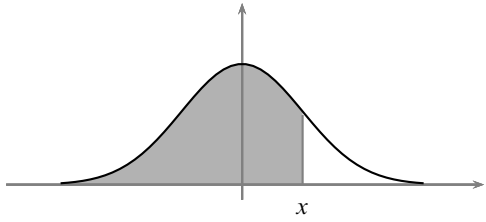
i) $\int u\,dv = uv - \int v\,du.$ (Integración por partes)

Fórmula de Stirling

$$n! \approx \sqrt{2\pi}\, n^{n+1/2}\, e^{-n}$$

n	$n!$	$Stirling$
1	1	0.922
2	2	1.919
3	6	5.836
4	24	23.506
5	120	118.019
6	720	710.078
7	5040	4980.395
8	40320	39902.395
⋮	⋮	⋮

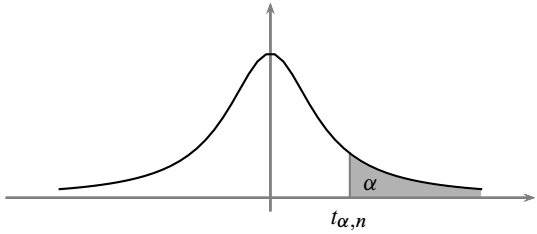
Tabla de la distribución normal estándar



$$\Phi(x) = P(X \leq x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt$$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8399
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

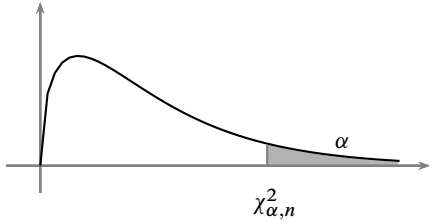
Tabla de la distribución $t(n)$



$$P(X \geq t_{\alpha,n}) = \alpha$$

$n\backslash\alpha$	0.005	0.01	0.025	0.05	0.1	0.15	0.20	0.25	0.30	0.35
1	63.657	31.821	12.706	6.314	3.078	1.963	1.376	1.000	0.727	0.510
2	9.925	6.965	4.303	2.920	1.886	1.386	1.061	0.816	0.617	0.445
3	5.841	4.541	3.182	2.353	1.638	1.250	0.978	0.765	0.584	0.424
4	4.604	3.474	2.776	2.132	1.533	1.190	0.941	0.741	0.569	0.414
5	4.032	3.365	2.571	2.015	1.476	1.156	0.920	0.727	0.559	0.408
6	3.707	3.143	2.447	1.943	1.440	1.134	0.906	0.718	0.553	0.404
7	3.499	2.998	2.365	1.895	1.415	1.119	0.896	0.711	0.549	0.402
8	3.355	2.896	2.306	1.860	1.397	1.108	0.889	0.706	0.546	0.399
9	3.250	2.821	2.262	1.833	1.383	1.100	0.883	0.703	0.543	0.398
10	3.169	2.764	2.228	1.812	1.372	1.093	0.879	0.700	0.542	0.397
11	3.106	2.718	2.201	1.796	1.363	1.088	0.876	0.697	0.540	0.396
12	3.055	2.681	2.179	1.782	1.356	1.083	0.873	0.695	0.539	0.395
13	3.012	2.650	2.160	1.771	1.350	1.079	0.870	0.694	0.538	0.394
14	2.977	2.624	2.145	1.761	1.345	1.076	0.868	0.692	0.537	0.393
15	2.947	2.602	2.131	1.753	1.341	1.074	0.866	0.691	0.536	0.393
16	2.291	2.583	2.120	1.746	1.337	1.071	0.865	0.690	0.535	0.392
17	2.898	2.567	2.110	1.740	1.333	1.069	0.863	0.689	0.534	0.392
18	2.878	2.552	2.101	1.734	1.330	1.067	0.862	0.688	0.534	0.392
19	2.861	2.539	2.093	1.729	1.328	1.066	0.861	0.688	0.533	0.391
20	2.845	2.528	2.086	1.725	1.325	1.064	0.860	0.687	0.533	0.391
21	2.831	2.518	2.080	1.721	1.323	1.063	0.859	0.686	0.532	0.391
22	2.819	2.508	2.074	1.717	1.321	1.061	0.858	0.686	0.532	0.390
23	2.807	2.500	2.069	1.714	1.319	1.060	0.858	0.685	0.532	0.390
24	2.797	2.492	2.064	1.711	1.318	1.059	0.857	0.685	0.531	0.390
25	2.787	2.485	2.060	1.708	1.316	1.058	0.856	0.684	0.531	0.390
26	2.779	2.479	2.056	1.706	1.315	1.058	0.856	0.684	0.531	0.390
27	2.771	2.473	2.052	1.703	1.314	1.057	0.855	0.684	0.531	0.389
28	2.763	2.467	2.048	1.701	1.313	1.056	0.855	0.683	0.530	0.389
29	2.756	2.462	2.045	1.699	1.311	1.055	0.854	0.683	0.530	0.389
30	2.750	2.457	2.042	1.697	1.310	1.055	0.854	0.683	0.530	0.389
40	2.704	2.423	2.021	1.684	1.303	1.050	0.851	0.681	0.529	0.388
60	2.660	2.390	2.000	1.671	1.296	1.046	0.848	0.679	0.527	0.387
∞	2.576	2.326	1.960	1.645	1.282	1.036	0.842	0.674	0.524	0.385

Tabla de la distribución $\chi^2(n)$



$$P(X \geq \chi^2_{\alpha,n}) = \alpha$$

$n \backslash \alpha$	0.995	0.990	0.975	0.950	0.900	0.100	0.050	0.025	0.010	0.005
1	0.0	0.0	0.0	0.0	0.02	2.71	3.84	5.02	6.63	7.88
2	0.01	0.02	0.05	0.10	0.21	4.61	5.99	7.38	9.21	10.60
3	0.07	0.11	0.22	0.35	0.58	6.25	7.81	9.35	11.34	12.84
4	0.21	0.30	0.48	0.71	1.06	7.78	9.49	11.14	13.28	14.86
5	0.41	0.55	0.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	0.68	0.87	1.24	1.64	2.20	10.65	12.59	14.45	16.81	18.55
7	0.99	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.72	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.27	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.87	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.28	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	46.95
28	12.46	13.57	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.42	104.22
80	51.17	53.54	57.15	60.39	64.28	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	107.57	113.14	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	118.50	124.34	129.56	135.81	140.17

Bibliografía

[1] Alonso Reyes, M.d.P. y Flores Díaz, J.A., Estadística descriptiva para bachillerato, Serie: Temas de Matemáticas para Bachillerato, Instituto de Matemáticas, UNAM, 2004.

[2] Blake, I.F., An introduction to applied probability, John Wiley and Sons, 1979.

[3] Blomm, G., Holst, L., y Sandell, D., Problems and snapshots from the world of probability, Springer-Verlag, 1994.

[4] Devore, J., Probability and statistics for the engineering and the sciences, Duxbury Press, 2008.

[5] Garza, T., Elementos de cálculo de probabilidades, UNAM, 1983.

[6] Godfrey, M.G., Roebuck, E.M., y Sherlock, A.J., Concise statistics, Edward Arnold, 1988.

[7] Granville, W.A., Cálculo diferencial e integral, Limusa, 1982.

[8] Hernández Del Valle, A. y Hernández Lerma, O., Elementos de probabilidad y estadística, Textos, vol. 21, Aportaciones Mat. Soc. Mat. Mexicana, 2003.

[9] Hoel, P.G., Port, S.C., y Stone, C.J., Introduction to probability theory, Houghton Mifflin, 1971.

[10] Hoel, P.G., Port, S.C., y Stone, C.J., Introduction to statistical theory, Houghton Mifflin, 1971.

[11] Hogg, R.V. y Tanis, E.A., Probability and statistical inference, Pearson/Prentice Hal, 2006.

[12] Kolmogorov, A.N., Foundations of the theory of probability, Chelsea Publishing Company, 1950.

[13] Mendenhall, W. y Sincich, T., Probabilidad y estadística para ingeniería y ciencias, Prentice Hall, 1997.

[14] Miller, I. y Miller, M., John E. Freund's mathematical statistics, Prentice Hall, 1999.

[15] Mood, A.M., Graybill, F.A., y Boes, D.C., Introduction to the theory of statistics, McGraw Hill, 1983.

[16] Olofsson, P., Probability, statistics, and stochastic processes, Wiley, 2005.

[17] Perero, M., Historia e historias de las matemáticas, Grupo Editorial Iberoamérica, 1994.

[18] Ross, S.M., Simulation, Academic Press, 2006.

[19] Ross, S.M., A first course in probability, Prentice Hall, 2009.

[20] Ross, S.M., Introduction to probability and statistics for engineers and scientists, Academic Press, 2009.

[21] Székely, G.J., Paradoxes in probability theory and mathematical statistics, D. Reidel Publishing Company, 2001.

[22] Velasco, G. y Wisniewski, P., Probabilidad y estadística para ingeniería y ciencia, Thomson, 2001.

[23] Wisniewski, P.M. y Bali, G., Ejercicios y problemas de la teoría de las probabilidades, Trillas, 1998.

Índice analítico

Axioma, 11

Bimodal, 92

Cantidad pivotal, 119
Coeficiente
  binomial, 19
  de correlación, 109
  de correlación muestral, 112
  multinomial, 19
Combinaciones, 18
Conjunto
  -s ajenos, 6
  -s operaciones, 4
  potencia, 7
Conjunto de datos
  bimodal, 92
  multimodal, 92
Covarianza, 108
  muestral, 111
Cuantil, 47
  de la dist. normal estándar, 69

Datos
  agrupados, 96
De Morgan
  leyes de, 6
Densidad
  conjunta, 101
  marginal, 104
Desviación estándar, 44
  de un conjunto de datos, 93
  para datos agrupados, 98
Diagrama de árbol, 7
Diferencia simétrica, 5
Distribución
  arcoseno, 166
  Bernoulli, 50
  beta, 63
  binomial, 51
  binomial negativa, 56
  conjunta, 103
  de una v.a., 40
  Erlang, 63
  exponencial, 61
  F, 72
  gama, 62
  geométrica, 52
  hipergeométrica, 58
  ji-cuadrada, 69
  marginal, 106
  normal, 65
  normal estándar, 66
  Poisson, 54
  t, 70
  uniforme continua, 60
  uniforme discreta, 49
  Weibull, 64
Distribución de prob.
  de una v.a., 40

Ensayo
  Bernoulli, 50
Escala
  de intervalo, 90
  de razón, 91
  nominal, 90
  ordinal, 90
Espacio
  equiprobable, 9
  muestral, 3
Esperanza, 40
  de una función de una v.a., 42
  propiedades, 43
Estadística, 89, 111
  descriptiva, 89
  inferencial, 89
Estandarización, 67
Estimación
  por intervalos, 118
  puntual, 112
Estimador
  de máxima verosimilitud, 115
  insesgado, 118
  máximo verosímil, 115
  puntual, 113

sesgado, 118
sesgo de un, 118
Evento, 3
-s ajenos, 6
compuesto, 4
simple, 4
Experimento
aleatorio, 2
determinista, 2

Fórmula
-s para derivadas, 213
-s para exponentes, 211
-s para integrales, 213
-s para logaritmos, 212
-s para sumas, 212
de Stirling, 214
Función
beta, 64
de acumulación conjunta, 103
de densidad, 34
de dist. $\mathrm{N}(0, 1)$, 67
de distribución, 36, 39
de probabilidad, 33
de verosimilitud, 114
gama, 62
indicadora, 51
Función de densidad
conjunta, 101
marginal, 104
Función de distribución
bivariada, 103
conjunta, 103
de un vector, 103
marginal, 106
Función de probabilidad, 33
conjunta, 100

Gráfica de pastel, 93
Grado de confianza, 119

Histogramas, 94

Imagen inversa, 31
Independencia
de dos eventos, 23
de dos variables aleatorias, 106
de variables aleatorias, 107
de varios eventos, 24
Insesgamiento, 117
Intervalo
de confianza, 119
grado de confianza, 119
lim inferior, 119
lim superior, 119

Ley de los grandes números, 73
Leyes de De Morgan, 6

Método
de la función inversa, 80
de máxima verosimilitud, 114
de momentos, 113
Media, 40, 92
aritmética, 92
de un conjunto de datos, 92
muestral, 111
Mediana, 93
de un conjunto de datos, 92
de una v.a., 48
Medida de probabilidad, 12
Medidas de dispersión, 93
Moda
de un conjunto de datos, 92
moda
de clase, 98
Momento
-s, 45
-s centrales, 47
muestral, 113
poblacional, 113
Momentos
método de, 113
Muestra, 88
aleatoria, 111
Multimodal, 92

Nivel de significancia, 128
Notación , 69

Ordenaciones
con repetición, 16
sin repetición, 17

Pérdida de memoria
dist exponencial, 165
dist. geométrica, 162
Partición, 24
Permutaciones, 18
Pie chart, 93
Población, 88
Polya
urna de, 152
Postulado, 11

Potencia de un conjunto, 7
Principio de multiplicación, 16
Probabilidad, 12
  axiomática, 11
  clásica, 8
  condicional, 22
  conjunta, 100
  de Laplace, 9
  de un evento, 8
  frecuentista, 9, 73
  subjetiva, 11
Producto
  Cartesiano, 7
  regla del, 152
Prueba
  varianza, 140
Prueba de hipótesis, 127
  nivel de significancia, 128
  región crítica, 128

Rango
  de un conjunto de datos, 93
Región crítica, 128
  tamaño de la, 128
Región de rechazo, 126
Regla del producto, 152

Sesgo, 118
Simulación
  de v.a.s, 76
  de v.a.s continuas, 80
  de v.a.s discretas, 77
  método de aceptación o rechazo, 82
  método de la función inversa, 80
Stirling, 214
Suceso, 3

Teorema
  central del límite , 74
  de Bayes, 26
  de probabilidad total, 24
  del binomio, 19, 149
  del estadístico inconsciente, 42
  multinomial, 20
Triángulo de Pascal, 19

Urna de Polya, 152

Valor
  esperado, 40
  promedio, 40
Variable, 89
  cualitativa, 89
  cuantitativa, 89
Variable aleatoria, 29
  continua, 30
  discreta, 30
  distribución de una, 40
Varianza, 43
  de un conjunto de datos, 93
  muestral, 111
  para datos agrupados, 98
  propiedades, 44
  prueba de hipótesis, 140
Vector aleatorio, 99
  continuo, 99
  discreto, 99

CURSO ELEMENTAL DE PROBABILIDAD Y ESTADÍSTICA, de la colección Aportaciones Matemáticas, editada por la Universidad Nacional Autónoma de México en colaboración con la Sociedad Matemática Mexicana, se terminó de imprimir el 18 de febrero de 2013 en los talleres de Publidisa Mexicana, S.A. de C.V., Chabacano 69, Col. Asturias, México 06850, D.F. Se tiraron 345 ejemplares, en papel Bond blanco de 90 gramos, encuadernación rústica con acabado en plastificado brillante. Se utilizó en la composición la familia tipográfica Times 10 pt. Tipo de impresión digital. El cuidado de la edición estuvo a cargo de Nils Ackermann.

Instituto de Matemáticas de la UNAM

La misión básica del Instituto de Matemáticas de la UNAM es realizar investigación en matemáticas del más alto nivel y en una variedad de áreas lo más amplia posible con el objetivo de que las matemáticas en México lleguen a desempeñar el papel estratégico que de ellas exige la sociedad contemporánea. Como parte consustancial de lo anterior, y en sintonía con el compromiso de la UNAM con el desarrollo del país, esta misión implica hacer un fuerte énfasis en la formación de recursos humanos y en la enseñanza a todos los niveles, incluyendo la divulgación.

El Instituto ha logrado desarrollar esta amplia gama de actividades de manera integral. A lo largo de sus 70 años de historia ha sabido sembrar la semilla de las matemáticas en otras partes de la República. Su apoyo a la formación de otras instituciones de investigación y enseñanza ha sido siempre decidido y en concordancia con los principios de nuestra máxima casa de estudios. Tan sólo en lo que se refiere al ámbito de la UNAM, destacamos la creación de tres Unidades Foráneas: Cuernavaca, Juriquilla y Morelia (esta última "Centro de Ciencias Matemáticas" en 2011). Actualmente se exploran otras formas de descentralización y desarrollo regional en Oaxaca.

http://www.matem.unam.mx

Directorio Institucional del Instituto de Matemáticas 2010–2014:

Javier Bracho	•	director
Sergio Rajsbaum	•	secretario académico
Armando Ramírez	•	secretario administrativo

Sección de Publicaciones:

Se encarga de la edición, difusión y distribución de las publicaciones periódicas, series y textos que se realizan en el Instituto.

Laura Ortiz (responsable académico).
Gabriela Sanginés (responsable técnico).
Leonardo Espinosa (responsable técnico de edición, consultoría y revisión especializadas).
Celia Osorio (responsable de inventario, control de ventas y apoyo logístico).
Helena Lluis (responsable de convenios, corrección de estilo y lectura de pruebas especializadas).

Sección de Cómputo:

Ricardo Gómez, Federico Cázarez, Rubén Alfaro, Carlos Rivera.

Sección de Difusión y Divulgación:

Gabriela Artigas, Víctor Hugo Alcántara, Lissette Martínez, Paloma Zubieta.

Sociedad Matemática Mexicana

Junta Directiva de la Sociedad Matemática Mexicana 2012–2014:

Luis Montejano Peimbert	•	presidente
Judith Zubieta García	•	vicepresidente
Ricardo A. Cantoral Uriza	•	secretario general
José Carlos Gómez Larrañaga	•	tesorero
Luz de Teresa de Oteyza	•	secretaria de actas
Lorenzo Héctor Juárez Valencia	•	vocal
Gabriela Araujo Pardo	•	vocal

La Sociedad Matemática Mexicana fue fundada en el año de 1943. Entre sus tareas fundamentales destacan: Estimular y mantener el interés por la investigación matemática en México; publicar revistas científicas; contribuir al mejoramiento de los programas de estudio y la enseñanza de las matemáticas en el país; estimular la elaboración y publicación de libros de texto en matemáticas; organizar conferencias, reuniones, congresos y concursos de matemáticas; cooperar en la resolución de los problemas matemáticos que se presentan en el sector productivo del país, así como los que surgen en las investigaciones de otras ciencias; promover el convencimiento entre la sociedad mexicana de que las matemáticas son parte integral de la cultura básica y de las habilidades de uso cotidiano de la población. http://www.smm.org.mx

Instituciones patrocinadoras de la Sociedad Matemática Mexicana:

- Asociación Mexicana para la Innovación en Tecnología Educativa, A.C.
- *Benemérita Universidad Autónoma de Puebla*
- *Centro de Ciencias Matemáticas-UNAM, Morelia*
- *Centro de Ingeniería y Desarrollo Industrial, Querétaro*
- *Centro de Investigación Científica y de Educación Superior de Ensenada*
- *Centro de Investigación en Matemáticas, A. C.*
- Consejo Mexiquense de Ciencia y Tecnología
- Corporación Mexicana de Investigación en Materiales, S.A. de C.V., Saltillo
- *Departamento de Matemática Educativa-CINVESTAV*
- *Departamento de Matemáticas-CINVESTAV*
- *Departamento de Matemáticas-UAM Iztapalapa*
- *Escuela Superior de Física y Matemáticas-IPN*
- *Facultad de Ciencias-UNAM*
- *Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas-UNAM*
- *Instituto de Matemáticas-UNAM*
- *Instituto Potosino de Investigación Científica y Tecnológica, A.C.*
- *Instituto Tecnológico Autónomo de México*
- *Instituto Tecnológico Superior de Lerdo, Durango*
- *Instituto Tecnológico Superior de Puerto Vallarta*
- *Universidad Autónoma de Coahuila*
- *Universidad Autónoma del Estado de Hidalgo*
- *Universidad Autónoma de Yucatán*
- *Universidad de las Américas Puebla*
- *Universidad de Sonora*
- *Universidad del Papaloapan, Oaxaca*
- *Universidad Juárez Autónoma de Tabasco*
- *Universidad Veracruzana*

Se han distinguido con letra cursiva los Miembros Institucionales de la SMM.

PUBLICACIONES DEL INSTITUTO DE MATEMÁTICAS, UNAM:

MONOGRAFÍAS DEL INSTITUTO DE MATEMÁTICAS

1 *Localization in non commutative rings.* B. J. Mueller (1975).
2 *Teoría de los números algebraicos.* A. Díaz-Barriga, A. I. Ramírez-Galarza, F. Tomás (1975).
3 *Integrales de medida positiva.* G. Zubieta (1976).
4 *Rings with polynomial identities.* B. J. Mueller (1977).
5 *Grupos profinitos, grupos libres y productos libres.* L. Ribes (1977).
6 *Introducción a la teoría de las clases características en la geometría algebraica.* A. Holme (1978).
7 *Embeddings, projective invariants and classifications.* A. Holme (1979).
8 *The relative spectral sequence of Leray-Serre for fibrations pairs.* C. Prieto (1979).
9 *Caminatas aleatorias y movimiento browniano.* D. B. Hernández (1981).
10 *On polarized varieties.* T. Matsusaka (1981).
11 *Métodos diagramáticos en teoría de representaciones.* C. Cibils, F. Larrión, L. Salmerón (1982).
12 *Espacios simplécticos sobre β-anillos y anillos de Hermite, trasvecciones simplécticas.* E. Fernández Bermejo (1982).
13 *Abelian integrals.* G. Kempf (1983).
14 *On Selfinjective algebras of finite representation type.* J. Waschbüsch (1983).
15 *On ideal theory in Banach and topological algebras.* W. Zelazko (1984).
16 *Teoría general de procesos e integración estocástica.* T. Bojdecki (1985).
17 *La figura espectral de operadores.* C. Hernández Garciadiego, E. de Oteyza (1986).
18 *Teoría de punto fijo.* A. Dold. Traducción: C. Prieto. Vol. I, Vol. II, Vol. III (1986).
19 *Projective embeddings of algebraic varieties.* J. Roberts (1988).
20 *Teoría general de procesos e integración estocástica.* 2a. Edición. T. Bojdecki (1989).
21 *Análisis funcional I.* C. Bosch Giral, E. Fernández Bermejo (1989).
22 *Introducción a la topología de las variedades de dimensión infinita.* L. Montejano (1989).
23 *Introducción a la teoría de representaciones de álgebras.* R. Martínez-Villa (1990).

MEMORIAS DEL 50 ANIVERSARIO DEL INSTITUTO DE MATEMÁTICAS. 1942–1992.

MATEMÁTICAS EN LA UNAM. MEMORIAS DEL 60 ANIVERSARIO DEL INSTITUTO DE MATEMÁTICAS. (2003).

Topología algebraica. Un enfoque homotópico. M. Aguilar, S. Gitler, C. Prieto. Coedición de Instituto de Matemáticas, UNAM – McGraw-Hill Interamericana Editores (1998).

TEMAS DE MATEMÁTICAS PARA BACHILLERATO.

1 *Cuando cuentes cuántos...* H. A. Rincón Mejía 1a. Reimp. 2a. Ed. (2009).
2 *Sistemas de ecuaciones y de desigualdades.* A. I. Ramírez Galarza. 1a. Reimp. 1a. Ed. (2008).
3 *La historia de un empujón: un vistazo a las ecuaciones diferenciales ordinarias y a los sistemas dinámicos.* L. Ortiz Bobadilla y E. Rosales González. 2a. Reimp. 1a. Ed. (2011).
4 *Dos o tres trazos.* S. Cárdenas Rubio. 2a. Ed. (2008).
5 *Estadística descriptiva para bachillerato.* M.P. Alonso Reyes, J.A. Flores Díaz.1a. Ed. (2004).
6 *Relaciones de equivalencia.* M. Cruz Terán. 1a. Ed. (2006).
7 *Mosaicos.* L. Hidalgo. 1a. Ed. (2007).
8 *Funciones Circulares.* M. Cruz Terán. 1a. Ed. (2008).
9 *Cónicas, cuádricas y aplicaciones.* A. I. Ramírez Galarza. 1a. Ed. (por aparecer).

PUBLICACIONES DEL INSTITUTO DE MATEMÁTICAS, UNAM

Y DE LA SOCIEDAD MATEMÁTICA MEXICANA:

CUADERNOS DE OLIMPIADAS DE MATEMÁTICAS.

1 *Combinatoria.* M. L. Pérez-Seguí. 3a. Reimp. 3a. Ed. (2010).
2 *Principios de olimpiada.* A. Illanes Mejía. 6a. Reimp. 1a. Ed. (2011).
3 *Geometría.* R. Bulajich, J.A. Gómez Ortega. 9a. Reimp. 1a. Ed. (2012).
4 *Geometría. Ejercicios y problemas.* R. Bulajich, J.A. Gómez Ortega. 6a. Reimp. 1a. Ed. (2010).
5 *Teoría de números.* M. L. Pérez-Seguí. 6a. Reimp. 1a. Ed. (2011).
6 *Desigualdades.* R. Bulajich, J.A. Gómez Ortega, R. Valdez Delgado. 1a. Reimp. 4a. Ed. (2012).
6a *Inequalities.* R. Bulajich, J.A. Gómez Ortega. 1a. Ed. (2005). (Nota: Traduc. de "Desigualdades").
7 *Olimpiadas en SLP, elemental.* R. Bulajich, C. J. Rubio. 1a. Ed. (2011).
8 *Olimpiadas en SLP, avanzado.* R. Bulajich, C. J. Rubio. 1a. Ed. (2012). (Nota: 7 y 8 en 2008 se publicó: "Las olimpiadas de matemáticas en San Luis Potosí, 1987-2005").
9 *Matemáticas preolímpicas.* M. L. Pérez-Seguí. 3a. Reimp. 1a. Ed. (2012).
10 *Combinatoria para olimpiadas internacionales.* P. Soberón. 1a. Ed. (2010).
11 *Problemas avanzados de olimpiada.* A. Alberro, R. Bulajich, C. J. Rubio. 1a. Ed. (2010).
12 *Combinatoria avanzada.* M. L. Pérez-Seguí. 1a. Ed. (2010).
13 *Principio de las casillas.* J.A. Gómez Ortega, R. Valdez Delgado, R. Vázquez Padilla. 1a. Ed. (2011).

APORTACIONES MATEMÁTICAS

Serie: INVESTIGACIÓN

1 *Coloquio de sistemas dinámicos. Memorias. Guanajuato, México, 1983.* Editado por J. A. Seade, G. Sienra (1985).
2 *Categorical topology - The complete work of Graciela Salicrup.* Edited by H. Herrlich, C. Prieto (1988).
3 *Sistemas dinámicos holomorfos en superficies.* X. Gómez-Mont, L. Ortiz-Bobadilla. 2a. Ed. (2004).
4 *Simposio de probabilidad y procesos estocásticos. Memorias. Guanajuato, México 1988.* Editado por M. E. Caballero, L. G. Gorostiza (1989).
5 *Topics in algebraic geometry. Proceedings. Guanajuato, México 1989.* Editado por L. Brambila-Paz, X. Gómez-Mont (1992).
6 *Seminario internacional de álgebra y sus aplicaciones. Memorias. México 1991.* Editado por L. M. Tovar, C. Rentería, R. H. Villarreal (1992).
7 *II Simposio de probabilidad y procesos estocásticos. I Encuentro México-Chile de análisis estocástico. Memorias. Guanajuato, México 1992.* Editado por M. E. Caballero, L. G. Gorostiza (1992).
8 *Taller de geometría diferencial sobre espacios de geometrías. Memorias. Guanajuato, México 1992.* Editado por L. del Riego, C. T. J. Dodson (1992).
9 *Poblaciones aleatorias ramificadas y sus equilibrios.* A. Wakolbinger (1994).
10 *Una Introducción a la geometría computacional a través de los teoremas de la galería de arte.* V. Estivill-Castro (1994).
11 *III Simposio de probabilidad y procesos estocásticos. Memorias. Hermosillo, México 1994.* Editado por M. E. Caballero, L. G. Gorostiza (1994).
12 *IV Simposio de probabilidad y procesos estocásticos. Memorias. Guanajuato, México 1996.* Editado por L. G. Gorostiza, J. A. León, J. A. López-Mimbela (1996).
13 *Taller de variedades abelianas y funciones theta. Memorias. Morelia, Mich., México 1996.* Editado por R. Rodríguez, J. M. Muñoz Porras, S. Recillas (1998).

14 *Modelos estocásticos.* Editado por J. M. González Barrios, L. G. Gorostiza (1998).
15 *Inverse limits.* W. T. Ingram (2000).
16 *Modelos Estocásticos II.* Editado por D. Hernández, J.A. López-Mimbela, R. Quezada (2001).
17 *Topics in infinitely divisible distributions and Lévy processes.* A. Rocha-Arteaga, K. Sato (2003).
18 *Parametric Optimization and Related Topics VII.* Editado por J. Guddat, H. Th. Jongen, J.-J. Rückmann, M. Todorov (2004).
19 *Continuum Theory: in Honor of Professor David P. Bellamy on the occasion of this 60th Birthay.* Edited by I. W. Lewis, S. Macías, S. B. Nadler, Jr. (2007).
20 *Proceedings of the Fourteenth International Conference on Fibonacci Numbers and their Applications.* Editado por Florian Luca, Pantelimon Stănică (2011).

Serie: COMUNICACIONES

1 *Programa de investigación del XVIII congreso nacional de la Sociedad Matemática Mexicana. Memorias. Mérida, México 1984.* Editado por M. Clapp, J. A. Seade (1986).
2 *Teoremas límite de alta densidad para campos aleatorios ramificados.* B. Fernández (1986).
3 *Programa del XIX congreso nacional de la Sociedad Matemática Mexicana, Vol. I. Memorias. Guadalajara, México 1986.* Editado por J.A. de la Peña, C.Prieto, G.Valencia, L.Verde(1987).
4 *Programa del XIX congreso nacional de la Sociedad Matemática Mexicana, Vol.II. Memorias. Guadalajara, México 1986.* Editado por J.A. de la Peña,C.Prieto,G.Valencia, L.Verde(1987).
5 *Programa del XX congreso nacional de la Sociedad Matemática Mexicana. Memorias. Xalapa, México 1987.* Editado por M. A. Aguilar, L. Salmerón, C. Vargas (1988).
6 *XXI Congreso nacional de la Sociedad Matemática Mexicana. Memorias. Hermosillo, Sonora 1988.* Editado por F. Aranda, J. Bracho, A. Sánchez Valenzuela, A. Vargas (1989).
7 *Breve introducción a códigos detectores-correctores de error.* C. Rentería, H. Tapia, W. Y. Vélez (1990).
8 *XXII Congreso nacional de la Sociedad Matemática Mexicana. Memorias. Puebla, Puebla 1989.* Editado por P. Barrera, A. Illanes, F. O'Reilly, S. Recillas (1990).
9 *XXIII Congreso nacional de la Sociedad Matemática Mexicana. Memorias. Guanajuato, México 1990.* Editado por A. García-Máynez, L. G. Gorostiza, J. Ize, M. Mendoza (1991).
10 *La estructura de los dendroides suaves.* S. Macías Alvarez (1993).
11 *XXIV Congreso nacional de la Sociedad Matemática Mexicana. Memorias. Oaxtepec, Morelos 1991.* Editado por O. Hernández, L. Montejano, B. Rumbos, A. A. Wawrzyñzyk (1992).
12 *XXV Congreso nacional de la Sociedad Matemática Mexicana Vol. I. Memorias. Xalapa, Veracruz 1992.* Editado por F. Larrión, A. Olvera, V. Pérez-Abreu, E. Vallejo (1993).
13 *XXV Congreso nacional de la Sociedad Matemática Mexicana Vol. II. Memorias. Xalapa, Veracruz 1992.* Editado por F. Larrión, A. Olvera, V. Pérez-Abreu, E. Vallejo (1993).
14 *XXVI Congreso nacional de la Sociedad Matemática Mexicana. Memorias. Morelia, Mich. 1993.* Editado por M. E. Caballero, J. Delgado, A. G. Raggi, J. Rosenblueth (1994).
15 *XI Escuela Latinoamericana de Matemáticas. Memorias. UNAM, México, D. F.; CIMAT, Gto. 1993.* Editado por X. Gómez-Mont, J. A. de la Peña, J. A. Seade (1994).
16 *XXVII Congreso nacional de la Sociedad Matemática Mexicana. Memorias. Querétaro, Qro. 1994.* Editado por J. A. León, A. Nicolás, F. Ongay, A. Tamariz (1995).
17 *Grupo de estudio con la industria y cursos en matemáticas industriales. Memorias. Oaxaca, Oaxaca 1995.* Editado por A. Fitt, R. Martínez-Villa (1996).
18 *XXVIII Congreso nacional de la Sociedad Matemática Mexicana. Memorias. Colima, Colima 1995.* Editado por J.L. Morales Pérez, S. Pérez Esteva, F. Sánchez Bringas, G. Villa Salvador (1996).
19 *Problemas combinatorios sobre conjuntos finitos de puntos.* B. M. Ábrego Lerma (1997).

20 *XXIX Congreso nacional de la Sociedad Matemática Mexicana. Memorias. San Luis Potosí, SLP 1996.* Editado por F. Avila Murillo, R. Montes-de-Oca, R. del Río Castillo, J. Muciño-Raymundo (1997).

21 *IV Escuela de verano de geometría y sistemas dinámicos. Memorias. Cimat, Guanajuato 1997.* Editado por O. Calvo, R. Iturriaga (1998).

22 *XXX Congreso nacional de la Sociedad Matemática Mexicana. Memorias. Aguascalientes, Ags. 1997.* Editado por A. López Mimbela, M. Neumann, M. Rzedowski, M. Shapiro (1998).

23 *Segundo grupo de estudio con la industria y cursos en matemáticas industriales. Memorias. Cocoyoc, Mor., México. 1997.* Editado por A. Fitt, R. Martínez-Villa, H. Ockendon (1999).

24 *3rd. International conference on approximation and optimization in the Caribean. Proceedings. Puebla, México. 1995.* Editado por B. Bank, J. Bustamante, J. Guddat, M. A. Jiménez, H. Th. Jongen, W. Römisch (1998).

25 *XXXI Congreso nacional de la Sociedad Matemática Mexicana. Memorias. Hermosillo, Son. 1998.* Editado por P. Padilla, R. Quiroga, C. Signoret, A. Soriano (1999).

26 *Tendencias interdisciplinarias de las matemáticas.* Editado por S. Gitler, C. Prieto (2000).

27 *XXXII Congreso nacional de la Sociedad Matemática Mexicana. Memorias. Guadalajara, Jal. 1999.* Editado por L. Hernández Lamoneda, R. Quezada, J. Martínez Bernal, H. Sánchez Morgado (2000).

28 *Lecturas Básicas en Topología General.* Editado por L. M. Villegas, A. Sestier, J. Olivares (2000).

29 *XXXIII Congreso nacional de la Sociedad Matemática Mexicana. Memorias. Saltillo, Coah. 2000.* Editado por J. Alfaro, M. Eudave, J. González Espino-Barros, E. Pérez Chavela (2001).

30 *XXXIV Congreso Nacional de la Sociedad Matemática Mexicana. Memorias. Toluca, Méx. 2001.* Editado por G. Contreras, C. Rentería, E. R. Rodríguez, C. Villegas Blas (2002).

31 *Tópicos de Geometría Algebraica.* Editado por L. Brambila, P.L. del Angel, A. García Zamora, J. Muciño (2002).

32 *XXXV Congreso Nacional de la Sociedad Matemática Mexicana. Memorias. Durango, Dgo. 2002.* Editado por M. Aguilar, R. Quiroga (2003).

33 (Número cancelado.)

34 *XXXVI Congreso Nacional de la Sociedad Matemática Mexicana. Memorias. Pachuca, Hgo. 2003.* Editado por M. Aguilar, R. Quiroga (2004).

35 *Memorias de la Sociedad Matemática Mexicana.* Editado por M. Aguilar, R. Quiroga (2005).

36 *Memorias de la Sociedad Matemática Mexicana.* Editado por M. Aguilar, R. Quiroga (2006).

37 *Memorias de la Sociedad Matemática Mexicana.* Editado por M. Aguilar, L. Hernández Lamoneda (2007).

38 *Memorias de la Sociedad Matemática Mexicana.* Editado por M. Aguilar, L. Hernández Lamoneda (2008).

39 *Modelos en estadística y probabilidad.* Editado por J.M. González-Barrios, J.A. León, A. Pérez, L.A. Rincón, J. Villa (2008).

40 *Memorias de la Sociedad Matemática Mexicana.* Editado por M. Aguilar, L. Hernández Lamoneda (2009).

41 *Memorias de la Sociedad Matemática Mexicana.* Editado por M. Aguilar, L. Hernández Lamoneda (2010).

42 *Las matemáticas a través de los 50 años de la ESFM del IPN.* Editado por Lino Feliciano Reséndis, Luis Manuel Tovar (2011).

43 *Memorias de la Sociedad Matemática Mexicana.* Editado por M. Aguilar, L. Hernández Lamoneda (2011).

44 *Modelos en estadística y probabilidad II.* Editado por J.M. González-Barrios M., J.A. León Vázquez, J. Villa Morales (2011).

45 *Memorias de la Sociedad Matemática Mexicana.* Editado por M. Aguilar, L. Hernández Lamoneda (2012).

Serie: TEXTOS

1 *Introducción a la topología – Graciela Salicrup.* Editado por J. Rosenblueth, C. Prieto. Nivel medio. 1a. Reimpresión (1997).
2 *Procesos estocásticos.* C. Tudor. Nivel avanzado. 3a. Edición (2002).
3 *Lectures on continuous-time Markov control processes.* O. Hernández-Lerma. Nivel avanzado (1994).
4 *Un curso de lógica matemática.* C. R. Videla. Nivel avanzado (1995).
5 *Rudimentos de mansedumbre y salvajismo en teoría de representaciones.* F. Larrión, A. G. Raggi, L. Salmerón. Nivel avanzado (1995).
6 *Teoría general de procesos e integración estocástica.* T. Bojdecki. Nivel avanzado. 1a. Reimp. (2004).
7 *Intersection theory.* S. X. Descamps. Nivel avanzado (1996).
8 *Inverse problems.* H. W. Engl. Nivel avanzado (1996).
9 *El ABC de los splines.* P. Barrera, V. Hernández, C. Durán. Nivel elemental (1996).
10 *Lo antiguo y lo nuevo acerca de los conjuntos convexos.* H. Hadwiger. Traducción: L. Montejano. Nivel medio (1998).
11 *Matemáticas para las ciencias naturales.* J. L. Gutiérrez Sánchez, F. Sánchez Garduño. Niveles medio y avanzado (1998).
12 *Introducción a la teoría de redes.* M. C. Hernández Ayuso. Nivel medio 2a. Edición (2005).
13 *Teoría de conjuntos (una introducción).* F. Hernández Hernández. Nivel medio 3a. Ed. (2011).
14 *Lectures on quantum probability.* A. M. Chebotarev. Nivel avanzado (2000).
15 *Construcción de procesos autosimilares con variancia finita.* J. E. Figueroa López. Nivel avanzado (2000).
16 *Grupos algebraicos y teoría de invariantes.* C. Sancho de Salas. Nivel avanzado (2001).
17 *Cohomología de Galois de campos locales.* F. Zaldívar. Nivel avanzado (2001).
18 *Dimension Theory: An Introduction with Exercises.* S.B. Nadler, Jr. Nivel avanzado (2002).
19 *Cómputo Numérico con aritmética de punto flotante IEEE. Con un teorema, una regla empírica y ciento un ejercicios.* M. L. Overton. Traducción: A. Casares Maldonado. Nivel medio (2002) Coedición SIAM.
20 *Topología Diferencial.* V. Guillemin, A. Pollack. Traducción: O. Palmas Velasco. Nivel medio (2003).
21 *Elementos de Probabilidad y Estadística.* A. Hernández-del-Valle, O. Hernández-Lerma. Nivel elemental (2003).
22 *Topología General.* D. Hinrichsen, J. L. Fernández Muñiz, A. Fraguela Collar, Á. Álvarez Prieto. Nivel medio (2003).
23 *Introducción a los grupos topológicos de transformaciones.* S. de Neymet U. Con la colaboración de R. Jiménez B. Nivel avanzado (2005).
24 *Breviario de teoría analítica de los números.* E. P. Balanzario. Nivel medio 1a. Reimpresión (2009) Coedición Reverté.
25 *Cálculo de probabilidades.* F. M. Hernández Arellano. Nivel elemental (2003).
26 *Números primos y aplicaciones.* F. Luca. Nivel avanzado (2004).
27 *Historia y desarrollo de la teoría de los continuos indescomponibles.* F.L. Jones. Traducción: S. Macías. Nivel medio (2004).
28 *Hiperespacios de continuos.* A. Illanes. Nivel medio (2004).
29 *Cadenas de Markov.* M.E. Caballero, V.M. Rivero, G. Uribe Bravo, C. Velarde. Nivel medio 2a. Edición (2008).
30 *The fixed point property for continua.* S.B. Nadler, Jr. Nivel avanzado (2005).
31 *Invitación a la teoría de los continuos y sus hiperespacios.* Editado por R. Escobedo, S. Macías, H. Méndez. Nivel medio (2006).

32 *Introducción a la teoría de grupos.* F. Zaldívar. Nivel medio 1a. Reimpresión (2009). Coedición Reverté.

33 *Hyperspaces of sets. A text with research questions.* S.B. Nadler, Jr. Nivel avanzado (2006).

34 *Introducción a la optimización no lineal.* E. Accinelli. Nivel avanzado (2009). Coedición Reverté.

35 *Graphs, rings and polyhedra.* I. Gitler, R. H. Villarreal. Nivel avanzado (2011).

36 *Introducción a la topología de conjuntos.* A. García Máynez. Nivel elemental (2011).

37 *Elementos de topología general.* F. Casarrubias, Á. Tamariz. Nivel medio (2012).

38 *Cálculo.* H. Arizmendi, A. Carrillo, M. Lara. Nivel elemental (2012).

Alberto Barajas: Su oratoria, sus matemáticas y sus enseñanzas. Edición: V. Neumann-Lara, I. Puga, S. Macías (2010) 346 p. Contiene 2 dvd.

Contemporary Mathematics

260 *First summer school in analysis and mathematical physics. Cuernavaca, Morelos, 1998.* Editado por S. Pérez-Esteva, C. Villegas-Blas. Contemporary Mathematics No. 260. Coedición de Aportaciones Matemáticas – American Mathematical Society (2000).

289 *Second summer school in analysis and mathematical physics. Cuernavaca, Morelos, 2000.* Editado por S. Pérez-Esteva, C. Villegas-Blas. Contemporary Mathematics No. 289. Coedición de Aportaciones Matemáticas – American Mathematical Society (2001).

336 *Stochastic Models.* Editado por J.M. González-Barrios, J.A. León, A. Meda. Contemporary Mathematics No. 336. Coedición de Aportaciones Matemáticas – American Mathematical Society (2003).

340 *Spectral Theory of Schrödinger Operators. Lecture Notes from a Workshop on Schrödinger Operator Theory. IIMAS, UNAM, Mexico, 2001.* Editado por R. del Río, C. Villegas-Blas. Contemporary Mathematics No. 340 Coedición de Aportaciones Matemáticas – American Mathematical Society (2004).

341 *Topological algebras and their applications. Fourth International Conference on Topological Algebras and Their Applications. Oaxaca, Mexico, 2002.* Editado por H. Arizmendi, C. Bosch, L. Palacios. Contemporary Mathematics No. 341. Coedición de Aportaciones Matemáticas – American Mathematical Society (2004).

389 *Geometry and Dynamics International Conference in Honor of the 60th Anniversary of Alberto Verjovsky. January 6-11, 2003. Cuernavaca, Mexico.* Editado por J. Eells, E. Ghys, M. Lyubich, J. Palis, J. Seade. Contemporary Mathematics No. 389. Coedición de Aportaciones Matemáticas – American Mathematical Society (2005).

476 *Fourth summer school in analysis and mathematical physics. Topics in Spectral Theory and Quantum Mechanics. Cuernavaca, Morelos, 2005.* Editado por C. Villegas-Blas. Contemporary Mathematics No. 476. Coedición de Aportaciones Matemáticas – American Mathematical Society (2008).

Información y pedidos:

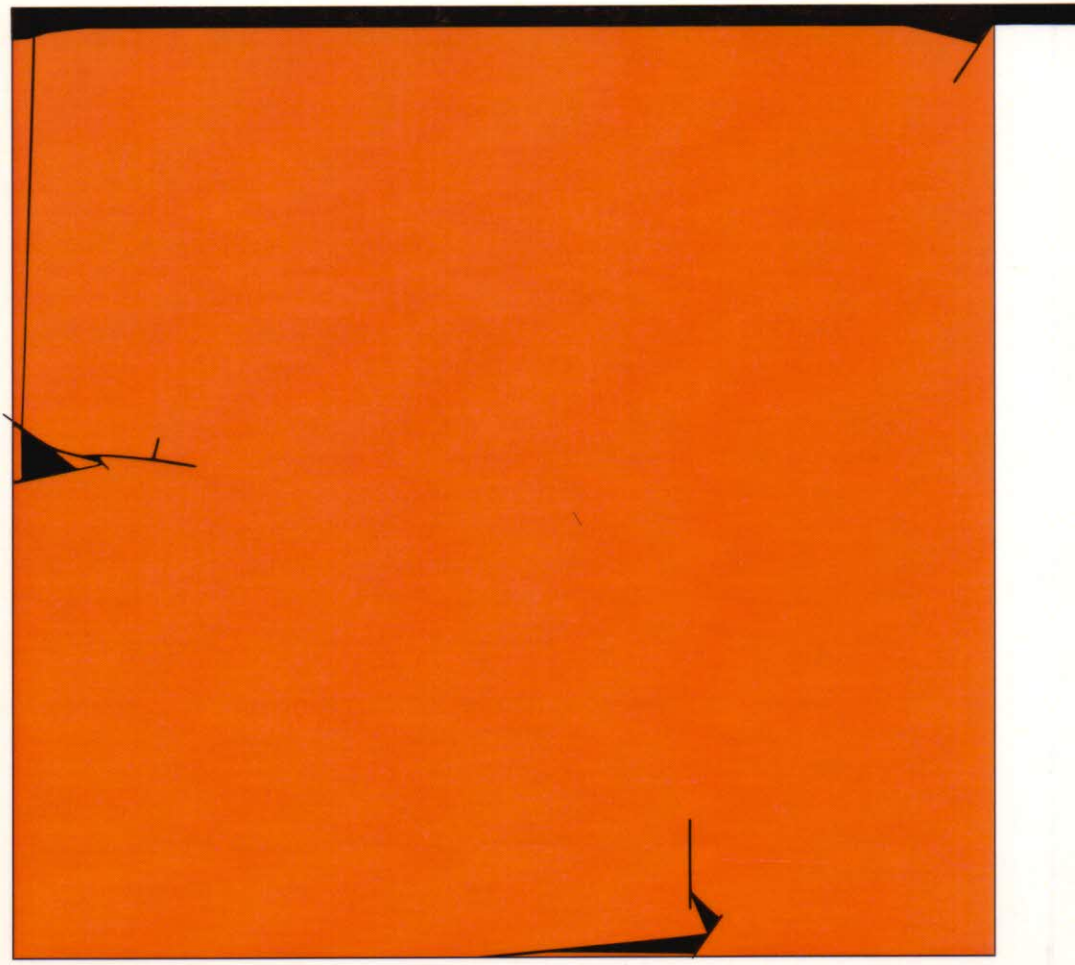
Gabriela Sanginés, Sección de Publicaciones
Instituto de Matemáticas, UNAM, Circuito Exterior
Ciudad Universitaria, 04510 México, D. F. MÉXICO
TEL: (52) (5) 622-44-96,
FAX: (52) (5) 550 13 42 y (52) (5) 616 03 48
e-mail: edicion@matem.unam.mx
WEB: http://www.matem.unam.mx
http://www.smm.org.mx

APORTACIONES MATEMÁTICAS es una publicación del Instituto de Matemáticas de la UNAM en colaboración con la Sociedad Matemática Mexicana. Su serie **TEXTOS** comprende libros de texto de matemáticas y sus aplicaciones, para la licenciatura y el posgrado en matemáticas y en otras disciplinas. Se divide en tres niveles:

- *Nivel elemental:*
  Dirigido a estudiantes de los primeros dos años de la licenciatura.
- *Nivel medio:*
  Dirigido a estudiantes de los dos últimos años de la licenciatura.
- *Nivel avanzado:* Dirigido a estudiantes de posgrado.

CURSO ELEMENTAL DE PROBABILIDAD Y ESTADÍSTICA, de Luis Rincón es un texto de *nivel elemental.*

En el presente texto el lector encontrará una exposición concisa de algunos conceptos y resultados elementales de la probabilidad y la estadística. Está dirigido a estudiantes de nivel superior, particularmente del área de ingeniería, quienes dentro de sus distintos ámbitos de especialización requieren conocer y saber aplicar algunas de las técnicas e ideas de estas dos disciplinas matemáticas.
En el texto se abordan los temas desde un enfoque práctico, haciendo mayor énfasis en el uso y aplicación de los resultados. Así, en ocasiones se omiten las demostraciones rigurosas matemáticas pero se exponen con cuidado y formalidad las hipótesis que llevan a las distintas conclusiones. A lo largo del texto se presentan numerosos ejemplos e ilustraciones para mostrar las ideas y las situaciones consideradas. Se incluye además una colección de alrededor de 300 ejercicios junto con sus soluciones o sugerencias de solución.

Diseño de portada: Rubén Valencia.

ISBN: 978-607-02-4092-8